STIC-Biotech/ChemLib

From: Rao, Manjunath N.
Sent: Thursday, June 28, 2001 7:49 AM
To: STIC-Biotech/ChemLib
Subject: sequence search request for 09/668,788

From: Manjunath N. Rao
Art Unit 1652, Room 10D04
Phone: 306-5681

Date: 6-28-01

Please search the following as soon as possible for application with serial number **09/668,788**

SEQ ID NO: 1 and 3 against all commercial nucleic acid databases including issued patents database and pending application database and provide **a print of all results.**

SEQ ID NO: 2 and 4 against all commercial protein databases including issued patents database and pending application database and provide **a print of all results.**

If you have any questions please call me at the above phone number.

Thanks

Manjunath N. Rao
Art Unit 1652
Room CM1, 10D04
Phone 306-5681

MANJUNATH N. RAO, Ph. D
Patent Examiner
AU 1652, CM1 10D04
703-306-5681

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 1E01 TEL: 308-3534

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 1E01 TEL: 308-3534

Access DB# 1

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner # :_____ Date: _____
Art Unit: _____ Phone Number 30_____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

_____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

****

STAFF USE ONLY

Searcher: Toby Port
Searcher Phone #: 308 3534
Searcher Location: _____
Date Searcher Picked Up: 6/29
Date Completed: 7/2
Searcher Prep & Review Time: 10
Clerical Prep Time: _____
Online Time: 10

Type of Search

NA Sequence (#) 2
AA Sequence (#) 2
Structure (#) _____
Bibliographic _____
Litigation _____
Fulltext _____
Patent Family _____
Other _____

Vendors and cost where applicable

STN _____
Dialog _____
Questel/Orbit _____
Dr.Link _____
Lexis/Nexis _____
Sequence Systems cg
WWW/Internet _____
Other (specify) _____

PTO-1590 (1-2000)

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 09:33:07 ; Search time 2892.01 Seconds
(without alignments)
6145.360 Million cell updates/sec

Title: US-09-668-788-1
Perfect score: 1149
Sequence: 1 ttgaataccaataaaagagt.....aagccaaagtgctatcgtaa 1149

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1149	100.0	1149	9	AX016296	AX016296 Sequence
2	c	1149	100.0	35040	2	BACYACA	L77246 Bacillus su
3		1149	100.0	213680	2	BSUB0012	Z99115 Bacillus su
4		135.4	11.8	3118	2	AF270166	AF270166 Staphyloc
5		135.4	11.8	3493	2	AF270394	AF270394 Staphyloc
6		120.4	10.5	7791	3	SAY14370	Y14370 Staphylococ
7		118	10.3	975	9	AX016297	AX016297 Sequence
8		69.6	6.1	292550	2	AP001513	AP001513 Bacillus

	9	53.8	4.7	13408	1	AE001958
c	10	49	4.3	7218	10	I66494
	11	48	4.2	1266	10	E14811
	12	48	4.2	2142	14	CSU62622
c	13	44.4	3.9	150583	88	AC026884
c	14	44.4	3.9	170934	79	AL354684
c	15	44.2	3.8	55593	2	BSAJ2571
c	16	44.2	3.8	208230	2	BSUB0008
c	17	44.2	3.8	216750	2	BSUB0007
c	18	43.2	3.8	1141	10	AX083744
	19	43.2	3.8	9070	13	AF148542
	20	43	3.7	1569	15	SOL249607
	21	42.8	3.7	2055	12	AB047476
	22	42.6	3.7	158293	85	AC004746
	23	42.6	3.7	172209	75	AC074253
	24	42.2	3.7	5742	96	PYYEL6
c	25	42	3.7	137641	78	AF273085
	26	41.8	3.6	104992	60	AC005504
c	27	41.8	3.6	150467	87	AC009541
	28	41.8	3.6	169546	60	AC004157
	29	41.8	3.6	172237	88	AC083869
	30	41.6	3.6	187610	67	AC022218
	31	41.4	3.6	5220	2	BBBRGABCD
c	32	41.4	3.6	30303	6	CELC18B2
	33	41.4	3.6	52971	1	AE001584
	34	41.4	3.6	157848	65	AC018455
c	35	41.4	3.6	167032	87	AC012157
	36	41.4	3.6	171255	86	AC007912
c	37	41.2	3.6	138890	86	AC007970
c	38	41.2	3.6	146000	91	CNS01RGD
c	39	41.2	3.6	151961	71	AC037443
c	40	41.2	3.6	161474	64	AC016198
c	41	41	3.6	49898	70	AC027136
c	42	41	3.6	127447	86	AC006344
c	43	41	3.6	190912	90	AL357559
c	44	41	3.6	198479	73	AC068483
c	45	40.8	3.6	140908	74	AC069516

AE001958 Deinococc
I66494 Sequence 14
E14811 cDNA encodi
U62622 Cucumis sat
AC026884 Homo sapi
AL354684 Homo sapi
AJ002571 Bacillus
Z99111 Bacillus su
Z99110 Bacillus su
AX083744 Sequence
AF148542 Arabidops
AJ249607 Spinacia
AB047476 Nicotiana
AC004746 Homo sapi
AC074253 Homo sapi
X55197 P.yoelii YE
AF273085 Homo sapi
AC005504 Plasmodiu
AC009541 Human Chr
AC004157 Plasmodiu
AC083869 Homo sapi
AC022218 Homo sapi
X87201 Borrelia bu
U40413 Caenorhabdi
AE001584 Borrelia
AC018455 Homo sapi
AC012157 Homo sapi
AC007912 Homo sapi
AC007970 Homo sapi
AL157915 Human chr
AC037443 Homo sapi
AC016198 Homo sapi
AC027136 Staphyloc
AC006344 Homo sapi
AL357559 Human DNA
AC068483 Homo sapi
AC069516 Homo sapi

ALIGNMENTS

```
RESULT   1
AX016296
LOCUS       AX016296     1149 bp    DNA             PAT       07-SEP-2000
DEFINITION  Sequence 1 from Patent WO9949052.
ACCESSION   AX016296
VERSION     AX016296.1  GI:10041859
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 1149)
  AUTHORS   Zaehringer,U., Heinz,E., Jorasch,P. and Wolter,F.P.
  TITLE     Processive glycosyltransferase
  JOURNAL   Patent: WO 9949052-A 1 30-SEP-1999;
            FORSCHUNGSZENTRUM BORSTEL (DE); ZAEHRINGER ULRICH (DE); HEINZ ERNST
            (DE); JORASCH PETRA (DE); WOLTER FRANK P (DE); GVS GES FUER ERWERB
            UND VERWER (DE)
FEATURES             Location/Qualifiers
     source          1. .1149
                     /organism="Bacillus subtilis"
                     /db_xref="taxon:1423"
BASE COUNT      380 a    217 c    244 g    308 t
ORIGIN
```

Query Match 100.0%; Score 1149; DB 9; Length 1149;
Best Local Similarity 100.0%; Pred. No. 3.3e-265;
Matches 1149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ttgaataccaataaaagagtattaattttgactgcaaattacggaaatggacatgtgcag 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTGAATACCAATAAAAGAGTATTAATTTTGACTGCAAATTACGGAAATGGACATGTGCAG 60

Qy     61 gtagccaaaacactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTAGCCAAAACACTTTATGAACAATGTGTACGGCTCGGCTTTCAGCATGTAACAGTTTCT 120

Qy    121 aatttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaa 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AATTTGTACCAAGAGTCAAATCCGATTGTTTCAGAGGTAACTCAATACCTTTATTTAAAA 180

Qy    181 agcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGCTTCTCAATCGGGAAACAGTTTTATCGTTTGTTTTATTACGGAGTTGACAAAATCTAT 240

Qy    241 aataaacgtaaattcaatatttacttttaaaatgggtaataaaagattgggcgaacttgtc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AATAAACGTAAATTCAATATTTACTTTAAAATGGGTAATAAAAGATTGGGCGAACTTGTC 300

Qy    301 gatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GATGAACATCAGCCCGATATTATTATTAATACATTTCCGATGATCGTCGTGCCGGAATAC 360

Qy    361 agacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGACGCCGAACTGGAAGAGTCATTCCTACCTTCAACGTTATGACTGATTTTTGTCTTCAT 420

Qy    421 aaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaa 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAAATTTGGGTTCACGAAAACGTGGATAAATATTATGTGGCGACAGATTACGTGAAGGAA 480

Qy    481 aaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcagg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AAACTGCTGGAGATCGGCACTCATCCAAGCAATGTAAAAATCACAGGAATTCCAATCAGG 540

Qy    541 ccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcacca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CCGCAATTTGAAGAATCCATGCCTGTTGGCCCGATATATAAAAAGTACAATCTTTCACCA 600

Qy    601 aacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AACAAAAAAGTGCTTCTGATCATGGCAGGTGCTCACGGTGTATTAAAGAACGTAAAAGAG 660

Qy    661 ctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CTGTGCGAAAACCTTGTCAAGGATGACCAAGTGCAAGTAGTTGTCGTGTGCGGGAAAAAT 720

Qy    721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ACGGCTTTAAAAGAATCTTTGAGTGCGCTTGAAGCGGAAAATGGTGACAAATTAAAAGTT 780

Qy    781 ctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgattaccaag 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CTGGGCTATGTGGAGCGCATTGATGAGCTATTTCGGATCACAGATTGCATGATTACCAAG 840

Qy    841 cccggcggcattactttgacagaagccacagccattggagtgcctgtcattctgtacaaa 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CCCGGCGGCATTACTTTGACAGAAGCCACAGCCATTGGAGTGCCTGTCATTCTGTACAAA 900

Qy    901 cccgtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcggagctgccatc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CCCGTGCCTGGCCAGGAAAAAGAAAATGCAAACTTCTTTGAAGACCGCGGAGCTGCCATC 960

Qy    961 gttgtgaaccgtcatgaagagattctcgagtcagtcacttcccttcttgcagatgaagat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GTTGTGAACCGTCATGAAGAGATTCTCGAGTCAGTCACTTCCCTTCTTGCAGATGAAGAT 1020

Qy   1021 accttgcatcgcatgaagaaaaacattaaggaccttcatttagcaaactcctctgaagtg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ACCTTGCATCGCATGAAGAAAAACATTAAGGACCTTCATTTAGCAAACTCCTCTGAAGTG 1080

Qy   1081 attttagaggatatcctgaaggaatcagaaatgatgaccgccaaacaaaaagccaaagtg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    1081 ATTTTAGAGGATATCCTGAAGGAATCAGAAATGATGACCGCCAAACAAAAAGCCAAAGTG 1140

Qy    1141 ctatcgtaa 1149
           |||||||||
Db    1141 CTATCGTAA 1149


RESULT   2
BACYACA/c
LOCUS       BACYACA     35040 bp    DNA             BCT       10-APR-1996
DEFINITION  Bacillus subtilis (YAC10-9 clone) DNA region between the serA and
            kdg loci.
ACCESSION   L77246
VERSION     L77246.1  GI:1256615
KEYWORDS    DNA-binding protein; adenine phosphoribosyltransferase; bcsA gene;
            bsaA gene; cold shock protein; cspD gene; degR gene; dfrA gene;
            dihydrofolate reductase; dihydroxyacid dehydratase;
            exodeoxyribonuclease; homoserine O-succinyltransferase; ilvA gene;
            ilvD gene; metB gene; naringenin-chalcone synthase; stress-related
            protein; threonine dehydratase; thyB gene; thymidylate synthase;
            transport protein; xpt gene; ypaQ gene; ypbQ gene; ypbR gene; ypbS
            gene; ypcP gene; ypdP gene; ypdQ gene; ypeP gene; ypeQ gene; ypfP
            gene; ypgQ gene; ypgR gene; yphP gene; ypiP gene; ypjP gene; ypjQ
            gene; ypkP gene; yplP gene; yplQ gene; ypmP gene; ypmQ gene; ypmR
            gene; ypmS gene; ypmT gene; ypnP gene; ypoP gene; yppP gene; yppQ
            gene; ypqP gene; ypwA gene.
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 35040)
  AUTHORS   Iwakura,M., Kawata,M., Tsuda,K. and Tanaka,T.
  TITLE     Nucleotide sequence of the thymidylate synthase B and dihydrofolate
            reductase genes contained in one Bacillus subtilis operon
  JOURNAL   Gene 64 (1), 9-20 (1988)
  MEDLINE   88284366
REFERENCE   2  (bases 1 to 35040)
  AUTHORS   Sorokin,A.V., Azevedo,V., Zumstein,E., Galleron,N., Ehrlich,S.D.
            and Serror,P.
  TITLE     Sequence analysis of the Bacillus subtilis chromosome region
            between the serA and kdg loci cloned in a yeast artificial
            chromosome
  JOURNAL   Microbiology (1996) In press
REFERENCE   3  (bases 1 to 35040)
  AUTHORS   Capuano,V., Galleron,N., Pujic,P., Sorokin,A.V. and Ehrlich,S.D.
  TITLE     Organization of the Bacillus subtilis chromosome between kdg and
            the attachment site of the SPbetta prophage - use of Long Accurate
            PCR and yeast artificial chromosomes for sequencing
  JOURNAL   Unpublished (1996)
FEATURES             Location/Qualifiers
     source          1. .35040
                     /organism="Bacillus subtilis"
                     /sub_strain="168"
                     /strain="Marburg"
                     /db_xref="taxon:1423"
                     /clone="YAC10-9"
                     /clone_lib="P. Serror"
                     /map="205 degrees"
     gene            1. .1095
                     /gene="ypwA"
     CDS             <1. .1095
                     /gene="ypwA"
                     /note="putative"
                     /citation=[2]
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA96610.1"
                     /db_xref="GI:1256616"
                     /translation="EFNKRFITYWGYQEHPYDALLDLFEPGVTVKVLDQLFAELKEAI
                     IPLVKQVTASGNKPDTSFITKAFPKEKQKELSLYFLQELGYDFDGGRLDETVHPFATT
                     LNRGDVRVTTRYDEKDFRTAIFGTIHECGHAIYEQNIDEALSGTNLSDGASMGIHESQ
                     SLFYENFIGRNKHFWTPYYKKIQEASPVQFKDISLDDFVRAINESKPSFIRVEADELT
                     YPLHIIIRYEIEKAIFSNEVSVEDLPSLWNQKYQDYLGITPQTDAEGILQDVHWAGGD
                     FGYFPSYALGYMYAAQLKQKMLEDLPEFDALLERGEFHPIKQWLTEKVHIHGKRKKPL
                     DIIKDATGEELNVRYLIDYLSNKYSNLYLL"
     terminator      1167. .1199
                     /note="putative"
                     /citation=[3]
     gene            1426. .2010
                     /gene="xpt"
     CDS             1426. .2010
                     /gene="xpt"
                     /EC_number="2.4.2.7"
                     /function="purine biosynthesis"
                     /note="27% identity with E.coli adenine
                     phosphoribosyltransferase"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="adenine phosphoribosyltransferase"
                     /protein_id="AAA96611.1"
                     /db_xref="GI:1256617"
                     /translation="MEALKRKIEEEGVVLSDQVLKVDSFLNHQIDPLLMQRIGDEFAS
                     RFAKDGITKIVTIESSGIAPAVMTGLKLGVPVVFARKHKSLTLTDNLLTASVYSFTKQ
                     TESQIAVSGTHLSDQDHVLIIDDFLANGQAAHGLVSIVKQAGASIAGIGIVIEKSFQP
                     GRDELVKLGYRVESLARIQSLEEGKVSFVQEVHS"
     gene            2007. .3323
                     /gene="ypaQ"
     CDS             2007. .3323
                     /gene="ypaQ"
                     /function="pyrimidine salvage"
                     /note="26% of identity to the Bacillus caldolyticus uracil
                     transport prot. . .; putative"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /product="transport protein"
                     /protein_id="AAA96612.1"
                     /db_xref="GI:1256618"
                     /translation="MRNGFGKTLSLGIQHVLAMYAGAIVVPLIVGKAMGLTVEQLTYL
                     VSIDIFMCGVATLLQVWSNRFFGIGLPVVLGCTFTAVSPMIAIGSEYGVSTVYGSIIA
                     SGILVILISFFFGKLVSFFPPVVTGSVVTIIGITLMPVAMNNMAGGEGSADFGDLSNL
                     ALAFTVLSIIVLLYRFTKGFIKSVSILIGILIGTFIAYFMGKVQFDNVSDAAVVQMIQ
                     PFYFGAPSFHAAPIITMSIVAIVSLVESTGVYFALGDLTNRRLTEIDLSKGYRAEGLA
                     VLLGGIFNAFPYTAFSQNVGLVQLTGIKKNAVIVVTGVILMAFGLFPKIAAFTTIIPS
                     AVLGGAMVAMFGMVIAYGIKMLSRIDFAKQENLLIVACSVGLGLGVTVVPDIFKQLPS
                     ALTLLTTNGIVAGSFTAVVLNIVYNVFSKAKKIEQEADLAEQKTAV"
     terminator      3322. .3341
                     /note="putative"
                     /citation=[3]
     gene            3397. .4494
                     /gene="bcsA"
     CDS             3397. .4494
                     /gene="bcsA"
                     /EC_number="2.3.1.74"
                     /note="25.6% of identity to the Lycopersicon esculentum
                     (tomato) naringenin-chalcone synthase; alternate gene name
                     ypbP; putative"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /product="naringenin-chalcone synthase"
                     /protein_id="AAA96613.1"
                     /db_xref="GI:1256619"
                     /translation="MAFILSIGTSLPAYNVNQEKAAEFARYMFQHSFKDIDRLLSSFK
                     NGQIHSRQFVKPIEWYKEGHSFEEKNQIYIEETLKHSRAAVRECLSHPEFFQEAIPYE
                     KVEAVFFVSSTGLSTPSIEARLMNELPFSPYTKRIPIWGLGCAGGASGLARAAEYCKA
                     YPEAFVLVISAELCSLTFQPEDKTKSNLIGTSLFGDGIAAALLCGEKADRRVSKLKLA
                     PKIMDAQSVLMKQSEDVMGWDFTDQGFKVIFSRDIPTLVEKWLKTNVQIFLDKHKLSF
                     HDISVFLAHPGGKKVIDAYIKSLGLSSEKLSSAQSILQKHGNMSSATILYVIKDHLQN
                     GHKKEAERGLIGALGPGFSSELLLFSWEKGA"
     gene            4498. .5004
                     /gene="ypbQ"
     CDS             4498. .5004
```

```
                     /gene="ypbQ"
                     /note="33.3% of identity in 60 aa with the Sacharomyces
                     cerevisiae farnesyl cysteine carboxyl-methyltransferase;
                     putative"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA96614.1"
                     /db_xref="GI:1256620"
                     /translation="MFWLLIAILIVQRAAEMAVARQNEQKVKKQGAIEFGESHYPYII
                     TMHILFFLSLIAEVLLMNKQPSSWWLGIAAVILSVQIVRYWALCSLGAYWNTKILVVP
                     GVELVKKGPYKWMKHPNYAVVILEILLIPLLYQAYVTMCLFSIFNAVLLSVRIRAEDK
                     ALQEYSVK"
     gene            5340. .8921
                     /gene="ypbR"
     CDS             5340. .8921
                     /gene="ypbR"
                     /note="26.7% of identity in 165 aa to a Thermophilic
                     bacterium hypothetical protein 6; putative"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA96615.1"
                     /db_xref="GI:1256621"
                     /translation="MTDQNRKELLHKTGELYKQFIENQDEQRAAKLAAVMKKAADEEV
                     YIAFTGHYSAGKSSLLNCLLMENILPTSPIPTSANLVVIRNGEKRVRLHTTDGACAEL
                     EGTYQKDKVQQYCKDGEQIESVEIFDRYTEIDSGVAYIDTPGIDSTDDAHFLSAASIL
                     HQADALFYVVHYNHVHAEENVKFLRSIKESIPNVYFIVNQIDRHDETETKFGDYQAQV
                     EEMLCNEGISREALYFTSVTEPDHPFNQMGALREELSRIEQQSKSNMQALTEQKVRNL
                     LKEHTEMLKKDETGAPSFAEQLNIHTGLVQSLRDQLDEAEKQMTEAEKRMQEEINRIL
                     KNANLTPFEMRELAAAFLESQEPSFKTGFFFSKAKTAQERDKRRNAFFSDVAKRTEAE
                     ADWHMIDTLHKLAKVFDVYTAESEKLIQAYRTPLDISIIEHAVKHGAAFSSEYVLQYT
                     KDLAELIRKEAKREAADIIKVLSAMVKERVSKDVQTINDRLVQESEKLVFLQEQARLE
                     NNAREKTDRLWAIWEEESACPMHIDTEWFKSKKTRVAAPEQKQGRSQLTAQPMPKSEI
                     KMEQEMPLQDQIKRFYTLSDILGECSMLLKQTSAFRERVKRLEERKFTLALFGGFSSG
                     KSSFANALVGERVLPSSPTPTTATINKITKPINGNLNKTANVVFKTEDDLTAEILQLT
                     GIPKEPAGRSFTEKWEKAVKKNRLQEEHVKLISNFLLAYEKYQQYIQEQKKLTIPLSE
                     LKPYVAEETTACAVKEVTVYYTCPLTEKGITIVDTPGASSMNKRHTELAFQYIKDADA
                     FFYMTYYQHSFSKGDRSFLRKLGLVKESLSMDKMFFIINAADLAKDKTELETVTDYVS
                     AELVKEGVYEPQLFTVSSKEELVGKPESFYNQFSKVRKHLDRFIEVDVKKASAAQLSS
                     EADKLCETVFQLHQSQHQSREEKEAQKQCLMLSFERTAADIEKRRNSKTIIEKVKKDT
                     REQLYHIAQRLSYFANDLLKSAFHPGLQNGDWKKNVSKAMTTALHEYLFEYIQEIKTL
                     DVRMSGFIERHINEEWLDHFQKTLNEDGYFSVYAGDQHSNGIQLKEVEPEIEERAFEQ
                     ELKEIKSPKQFFEQKGKATFIEAVRMKLTKITEAWIKNEEESLISHYTAHLRRLQEDM
                     GEKAIAQITDQKETYLRGYAEGEHAKEIEMAYQACISWKNSDNTIKM"
     gene            8986. .9243
                     /gene="ypbS"
     CDS             8986. .9243
                     /gene="ypbS"
                     /note="putative"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA96616.1"
                     /db_xref="GI:1256622"
                     /translation="MSEVHKAISAHSSKQHEHIKAFMRLENMRELAIEEAVAKCRNDE
                     PYTTDAINEITEQMNQLAKKGIVPTRRLVSKEMVREYVSRM"
     terminator      9474. .9505
                     /note="putative"
                     /citation=[3]
     gene            9565. .10455
                     /gene="ypcP"
     CDS             9565. .10455

  Query Match             100.0%;  Score 1149;  DB 2;  Length 35040;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-265;
  Matches 1149;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ttgaataccaataaaagagtattaattttgactgcaaattacggaaatggacatgtgcag 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14936 TTGAATACCAATAAAAGAGTATTAATTTTGACTGCAAATTACGGAAATGGACATGTGCAG 14877

Qy      61 gtagccaaaacactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14876 GTAGCCAAAACACTTTATGAACAATGTGTACGGCTCGGCTTTCAGCATGTAACAGTTTCT 14817

Qy     121 aatttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaa 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14816 AATTTGTACCAAGAGTCAAATCCGATTGTTTCAGAGGTAACTCAATACCTTTATTTAAAA 14757

Qy     181 agcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctat 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14756 AGCTTCTCAATCGGGAAACAGTTTTATCGTTTGTTTTATTACGGAGTTGACAAAATCTAT 14697

Qy     241 aataaacgtaaattcaatatttactttaaaatgggtaataaaagattgggcgaacttgtc 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14696 AATAAACGTAAATTCAATATTTACTTTAAAATGGGTAATAAAAGATTGGGCGAACTTGTC 14637

Qy     301 gatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatac 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14636 GATGAACATCAGCCCGATATTATTATTAATACATTTCCGATGATCGTCGTGCCGGAATAC 14577

Qy     361 agacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcat 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14576 AGACGCCGAACTGGAAGAGTCATTCCTACCTTCAACGTTATGACTGATTTTTGTCTTCAT 14517

Qy     421 aaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaa 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14516 AAAATTTGGGTTCACGAAAACGTGGATAAATATTATGTGGCGACAGATTACGTGAAGGAA 14457

Qy     481 aaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcagg 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14456 AAACTGCTGGAGATCGGCACTCATCCAAGCAATGTAAAAATCACAGGAATTCCAATCAGG 14397

Qy     541 ccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcacca 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14396 CCGCAATTTGAAGAATCCATGCCTGTTGGCCCGATATATAAAAAGTACAATCTTTCACCA 14337

Qy     601 aacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagag 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14336 AACAAAAAAGTGCTTCTGATCATGGCAGGTGCTCACGGTGTATTAAAGAACGTAAAAGAG 14277

Qy     661 ctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14276 CTGTGCGAAAACCTTGTCAAGGATGACCAAGTGCAAGTAGTTGTCGTGTGCGGGAAAAAT 14217

Qy     721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14216 ACGGCTTTAAAAGAATCTTTGAGTGCGCTTGAAGCGGAAAATGGTGACAAATTAAAAGTT 14157

Qy     781 ctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgattaccaag 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14156 CTGGGCTATGTGGAGCGCATTGATGAGCTATTTCGGATCACAGATTGCATGATTACCAAG 14097

Qy     841 cccggcggcattactttgacagaagccacagccattggagtgcctgtcattctgtacaaa 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14096 CCCGGCGGCATTACTTTGACAGAAGCCACAGCCATTGGAGTGCCTGTCATTCTGTACAAA 14037

Qy     901 cccgtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcggagctgccatc 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14036 CCCGTGCCTGGCCAGGAAAAAGAAAATGCAAACTTCTTTGAAGACCGCGGAGCTGCCATC 13977

Qy     961 gttgtgaaccgtcatgaagagattctcgagtcagtcacttcccttcttgcagatgaagat 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13976 GTTGTGAACCGTCATGAAGAGATTCTCGAGTCAGTCACTTCCCTTCTTGCAGATGAAGAT 13917

Qy    1021 accttgcatcgcatgaagaaaaacattaaggaccttcatttagcaaactcctctgaagtg 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13916 ACCTTGCATCGCATGAAGAAAAACATTAAGGACCTTCATTTAGCAAACTCCTCTGAAGTG 13857

Qy    1081 attttagaggatatcctgaaggaatcagaaatgatgaccgccaaacaaaaagccaaagtg 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13856 ATTTTAGAGGATATCCTGAAGGAATCAGAAATGATGACCGCCAAACAAAAAGCCAAAGTG 13797

Qy    1141 ctatcgtaa 1149
           |||||||||
```

```
Db  13796 CTATCGTAA 13788


RESULT   3
BSUB0012
LOCUS       BSUB0012   213680 bp    DNA             BCT       26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 12 of 21): from 2195541
            to 2409220.
ACCESSION   Z99115 AL009126
VERSION     Z99115.1  GI:2634478
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 213680)
  AUTHORS   Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
            Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
            Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
            Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
            Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
            Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
            Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
            Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
            Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
            Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
            Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
            Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
            Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
            Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
            Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
            Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
            Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
            Danchin,A.
  TITLE     The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
  JOURNAL   Nature 390 (6657), 249-256 (1997)
  MEDLINE   98044033
REFERENCE   2  (bases 1 to 213680)
  AUTHORS   Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
            Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
            Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
            adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
            68 89 48
FEATURES             Location/Qualifiers
     source          1. .213680
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            complement(90. .782)
                     /gene="yoqL"
     CDS             complement(90. .782)
                     /gene="yoqL"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13977.1"
                     /db_xref="GI:2634479"
                     /db_xref="SPTREMBL:O34581"
                     /translation="MDSKKHKIINGYYHKNCISCKSWLPATEENFYSVKKNKDGLHSY
                     CKACVLKKAKESMLKNYDKQLERMRERNLLPGMKEAKKKYNSSLKKKKTQQIWQEKNK
                     LKLKNYRLQRDAHKKHNITDVQWQKCKDYFNNKCSYCGLKIEDHKILFKGTYIQSDFH
                     KEHVDHKGANDISNCIPACKSCNSSKHDFAFEEWYNSSNKNFSSERLLKIKEWLNRFK
                     EERQDSEWRTKG"
     gene            complement(822. .1025)
                     /gene="yoqK"
     CDS             complement(822. .1025)
                     /gene="yoqK"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13978.1"
                     /db_xref="GI:2634480"
                     /db_xref="SPTREMBL:O34326"
                     /translation="MKLIEERTYEERLYIYSDDIEARKHFVSELKPKGWMVDNCWVGI
                     DLNQIKQFYRFKRELTDDYVFNR"
     gene            complement(1045. .1560)
                     /gene="yoqJ"
     CDS             complement(1045. .1560)
                     /gene="yoqJ"
                     /function="unknown"
                     /note="similar to hypothetical proteins from B. subtilis"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13979.1"
                     /db_xref="GI:2634481"
                     /db_xref="SPTREMBL:O34359"
                     /translation="MKNPTMLKLKDKLLAVIEELITKENKYRFITGGALGTDQAACWC
                     VHILKKKHPHIKNIIATPFKEQDKVWSADQKMWYKRMLNVADEIINVEELDKYKVSGE
                     KPGEFSPAKMQKRNEYMIDHSEAIVAVYDGSKSGTRNCLNYAKKTYLGHQLWRLHPDF
                     NFELDITYFVG"
     gene            complement(1771. .1965)
                     /gene="yoqI"
     CDS             complement(1771. .1965)
                     /gene="yoqI"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13980.1"
                     /db_xref="GI:2634482"
                     /db_xref="SPTREMBL:O34339"
                     /translation="MIKSIILPEENTKITVGKPINDESNTKVIAIYDYREEPEEAFWV
                     HLSNGNDLFVDNHEVIVEYE"
     gene            complement(2014. .2466)
                     /gene="yoqH"
     CDS             complement(2014. .2466)
                     /gene="yoqH"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13981.1"
                     /db_xref="GI:2634483"
                     /db_xref="SPTREMBL:O34999"
                     /translation="MKRFILVLSFLSIIVAYPIQTNASPMPCSVILEPVDKNLKNAKG
                     VALIYKVQLNPPSAARTNISILAVHLPAPSSFGNYDSYEGFATKPGEISWRFKLYPTP
                     EEESPSWAGRIDTISAEMKNVKVQVRLSNSSTQKLGPSILTKNIESCY"
     gene            complement(2549. .2806)
                     /gene="yoqG"
     CDS             complement(2549. .2806)
                     /gene="yoqG"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13982.1"
                     /db_xref="GI:2634484"
                     /db_xref="SPTREMBL:O35030"
                     /translation="MEKMNLLKEITIFDLNKIKPGTKVQVTWYKGTEMEYTHNGEVII
                     NNGEKFYYNYVDKEGYVGHCHVNALDLKNYPDSLIVEIKSK"
     gene            complement(2851. .3054)
                     /gene="yoqF"
     CDS             complement(2851. .3054)
                     /gene="yoqF"
```

```
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13983.1"
                     /db_xref="GI:2634485"
                     /db_xref="SPTREMBL:O34584"
                     /translation="MKFAPMDKVKFKTASHLNKLRTLKKRVPELDDPLLGECWEFEED
                     GLRQFDWEENYEFVARPKHFNWD"
     gene            complement(3063. .3227)
                     /gene="yoqE"
     CDS             complement(3063. .3227)
                     /gene="yoqE"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13984.1"
                     /db_xref="GI:2634486"
                     /db_xref="SPTREMBL:O34429"
                     /translation="MENNVLYGVYSTRSRKFCFGIEEPSKTKARKELFNRIGTDAYKW
                     RFEIRKIKRK"
     gene            complement(3281. .4036)
                     /gene="yoqD"
     CDS             complement(3281. .4036)
                     /gene="yoqD"
                     /function="unknown"
                     /note="similar to phage-related DNA-binding protein
                     anti-repressor"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13985.1"
                     /db_xref="GI:2634487"
                     /db_xref="SPTREMBL:O34449"
                     /translation="MESYLTVIEQNGQLLVDSREVAEMVGKRHTDLLRSIDGYVAILL
                     NAKLRSVEFFLESTYKDATGRSLKHFHLTRKGCDMVANKMTGAKGVLFTAQYVSKFEE
                     MEKALKARPSLIDTYLDMNEDERAIAYFTERKAKRELQEQLTLAEPKVEKYDRFLNTD
                     GLMKIGQVAKAIGIKGMGQNNLFRFLRENKVLIDGTNKNAPYQKYVERGFFQVKTQET
                     SVGIKTITLVTPKGADFIVDLLKKHGHKREIAS"
     gene            complement(4077. .4484)
                     /gene="yoqC"
     CDS             complement(4077. .4484)
                     /gene="yoqC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13986.1"
                     /db_xref="GI:2634488"
                     /db_xref="SPTREMBL:O34834"
                     /translation="MEVNVKTNQREKFIRNGIPYDELDTQMIHLIDILNFKIGLKTRH
                     CCFGHKPYEEIQVMFEDEVNIKEDQILELAELAGREWKGLQLSFSKWARFSPLMFNWS
                     LVLSKRFRNPEDPNKYRYLRSVEEFFESYAAKK"
     gene            complement(4491. .4829)
                     /gene="yoqB"
     CDS             complement(4491. .4829)
                     /gene="yoqB"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13987.1"
                     /db_xref="GI:2634489"

  Query Match             100.0%;  Score 1149;  DB 2;  Length 213680;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-265;
  Matches 1149;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ttgaataccaataaaagagtattaatttgactgcaaattacggaaatggacatgtgcag 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110216 TTGAATACCAATAAAAGAGTATTAATTTGACTGCAAATTACGGAAATGGACATGTGCAG 110275

Qy     61 gtagccaaaacactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110276 GTAGCCAAAACACTTTATGAACAATGTGTACGGCTCGGCTTTCAGCATGTAACAGTTTCT 110335

Qy    121 aatttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaa 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110336 AATTTGTACCAAGAGTCAAATCCGATTGTTTCAGAGGTAACTCAATACCTTTATTTAAAA 110395

Qy    181 agcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110396 AGCTTCTCAATCGGGAAACAGTTTTATCGTTTGTTTTATTACGGAGTTGACAAAATCTAT 110455

Qy    241 aataaacgtaaattcaatatttacttttaaaatgggtaataaaagattgggcgaacttgtc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110456 AATAAACGTAAATTCAATATTTACTTTAAAATGGGTAATAAAAGATTGGGCGAACTTGTC 110515

Qy    301 gatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110516 GATGAACATCAGCCCGATATTATTATTAATACATTTCCGATGATCGTCGTGCCGGAATAC 110575

Qy    361 agacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110576 AGACGCCGAACTGGAAGAGTCATTCCTACCTTCAACGTTATGACTGATTTTTGTCTTCAT 110635

Qy    421 aaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaa 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110636 AAAATTTGGGTTCACGAAAACGTGGATAAATATTATGTGGCGACAGATTACGTGAAGGAA 110695

Qy    481 aaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcagg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110696 AAACTGCTGGAGATCGGCACTCATCCAAGCAATGTAAAAATCACAGGAATTCCAATCAGG 110755

Qy    541 ccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcacca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110756 CCGCAATTTGAAGAATCCATGCCTGTTGGCCCGATATATAAAAAGTACAATCTTTCACCA 110815

Qy    601 aacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110816 AACAAAAAAGTGCTTCTGATCATGGCAGGTGCTCACGGTGTATTAAAGAACGTAAAAGAG 110875

Qy    661 ctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110876 CTGTGCGAAAACCTTGTCAAGGATGACCAAGTGCAAGTAGTTGTCGTGTGCGGGAAAAAT 110935

Qy    721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110936 ACGGCTTTAAAAGAATCTTTGAGTGCGCTTGAAGCGGAAAATGGTGACAAATTAAAAGTT 110995

Qy    781 ctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgattaccaag 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 110996 CTGGGCTATGTGGAGCGCATTGATGAGCTATTTCGGATCACAGATTGCATGATTACCAAG 111055

Qy    841 cccggcggcattactttgacagaagccacagccattggagtgcctgtcattctgtacaaa 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111056 CCCGGCGGCATTACTTTGACAGAAGCCACAGCCATTGGAGTGCCTGTCATTCTGTACAAA 111115

Qy    901 cccgtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcggagctgccatc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111116 CCCGTGCCTGGCCAGGAAAAAGAAAATGCAAACTTCTTTGAAGACCGCGGAGCTGCCATC 111175

Qy    961 gttgtgaaccgtcatgaagagattctcgagtcagtcacttcccttcttgcagatgaagat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111176 GTTGTGAACCGTCATGAAGAGATTCTCGAGTCAGTCACTTCCCTTCTTGCAGATGAAGAT 111235

Qy   1021 accttgcatcgcatgaagaaaaacattaaggaccttcatttagcaaactcctctgaagtg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111236 ACCTTGCATCGCATGAAGAAAAACATTAAGGACCTTCATTTAGCAAACTCCTCTGAAGTG 111295

Qy   1081 attttagaggatatcctgaaggaatcagaaatgatgaccgccaaacaaaaagccaaagtg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 111296 ATTTTAGAGGATATCCTGAAGGAATCAGAAATGATGACCGCCAAACAAAAAGCCAAAGTG 111355

Qy   1141 ctatcgtaa 1149
          |||||||||
Db 111356 CTATCGTAA 111364


RESULT   4
```

```
AF270166
LOCUS       AF270166     3118 bp    DNA             BCT       01-AUG-2000
DEFINITION  Staphylococcus epidermidis strain SR1 clone step.1051e11 genomic
            sequence.
ACCESSION   AF270166
VERSION     AF270166.1  GI:9624074
KEYWORDS    .
SOURCE      Staphylococcus epidermidis.
  ORGANISM  Staphylococcus epidermidis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE   1  (bases 1 to 3118)
  AUTHORS   Kimmerly,W.J., Taylor,J.David, Nelsen,A.J., Godlevski,M.M.,
            Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
            Listenbee,S., Ashanti,C., Altshuller,G., Mamo,L., Shepherd,N.S.,
            Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
            Furdon,P.J.
  TITLE     Transposon-mediated sequencing of the Staphylococcus epidermidis
            genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3118)
  AUTHORS   Taylor,J.David, Kimmerly,W.J., Nelsen,A.J., Godlevski,M.M.,
            Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
            Listenbee,S., Ashanti,C., Altshuller,G., Mamo,L., Shepherd,N.S.,
            Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
            Furdon,P.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2000) Departments of Genomic Sciences and
            Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
            Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES             Location/Qualifiers
     source          1. .3118
                     /organism="Staphylococcus epidermidis"
                     /strain="SR1"
                     /db_xref="taxon:1282"
                     /clone="step.1051e11"
BASE COUNT     1048 a    480 c    492 g   1098 t
ORIGIN
```

Query Match 11.8%; Score 135.4; DB 2; Length 3118;
Best Local Similarity 46.7%; Pred. No. 4.3e-22;
Matches 510; Conservative 0; Mismatches 566; Indels 15; Gaps 2;

```
Qy     10 aataaaagagtattaattttgactgcaaattacggaaatggacatgtgcaggtagccaaa 69
Db    287 AATAAAAAGATATTGATTATTACTGGATCATTTGGTAACGGTCATATGCAAGTCACGCAA 346

Qy     70 acactttatgaacaatgtgtacggctcggctttcagcatgtaacagt------ttctaat 123
Db    347 AGTATTGTCAACCAATTGAATGAGATGAATCTCAATCATTTATCAGTCATTCAACATGAT 406

Qy    124 ttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaaagc 183
Db    407 TTGTTTATGGAAGCTCATCCAATTATGACTTCTATATGTAAGAAATGGTATATCAATAGC 466

Qy    184 ttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctataat 243
Db    467 TTTAAATATTTTAGAAATACATATAAACGATTTTACTATAGTCGCCCTAATGAGCTCGAT 526

Qy    244 aaacgtaaattcaatatttactttaaaatgggtaataaaagattgggcgaacttgtcgat 303
Db    527 AAATGTTTTTATAAATATTA---------TGGATTAAATAAACTCATCAACTTACTTATT 577

Qy    304 gaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatacaga 363
Db    578 AAAGAAAAGCCTGATCTCATATTATTAACATTTCCAACACCTGTGATGTCAGTGTTGACC 637

Qy    364 cgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcataaa 423
Db    638 GAACAATTTAATATAAATATCCCTATTGCGACAGTTATGACAGATTATCGCATGCATAAA 697

Qy    424 atttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaaaaa 483
Db    698 AATTGGCATTACACCATATTCACAAAGATATTATGTAGCAACAAAAGATACTAAAGATGAT 757

Qy    484 ctgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcaggccg 543
Db    758 TTCATTGAAGCTGGTGTTCCTGCTTCATATATTAAAGTGACGGGCATTCCTATTGCTGAT 817

Qy    544 caatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcaccaaac 603
Db    818 AAATTTGAAGAATCTATTGATAAAGAAGAATGGTTATCGCAACAACATTTAGACCCTTCA 877

Qy    604 aaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagagctg 663
Db    878 AAACCTACTATATTAATGTCAGCAGGTGCATTTGGTGTTTCAAAAGGCTTTGACTATATG 937

Qy    664 tgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaatacg 723
Db    938 ATTAATAATATTTTAGAAAAAAGTCCAAATTCGCAAGTGGTCATGATTTGTGGACGTAGT 997

Qy    724 gctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagttctg 783
Db    998 AAGGAACTTAAACGTTCATTAAAAGCTAAGTTCAAAGATAATCCAAGTGTAATAATATTA 1057

Qy    784 ggctatgtggagcgcattgatgagctatttcggatcacagattgcatgattaccaagccc 843
Db   1058 GGATATACAAATCACATGAATGAGTGGATGGCATCAAGCCAACTAATGATTACAAAACCT 1117

Qy    844 ggcggcattactttgacagaagccacagccattggagtgcctgtcattctgtacaaaccc 903
Db   1118 GGTGGTATCACAATTTCCGAAGGACTTAGTCGTTGTATTCCTATGATTTTTTTAAACCCT 1177

Qy    904 gtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcggagctgccatcgtt 963
Db   1178 GCACCCGGTCAAGAACTTGAAAATGCATATTACTTTGAAAGTAAAGGATTTGGAAAAATA 1237

Qy    964 gtgaaccgtcatgaagagattctcgagtcagtcacttcccttcttgcagatgaagatacc 1023
Db   1238 GCAGATACTCCAAATGAGGCAATTGATATTGTTTCTGACTTAACAAATAACGAAGAGACT 1297

Qy   1024 ttgcatcgcatgaagaaaaacattaaggaccttcatttagcaaactcctctgaagtgatt 1083
Db   1298 TTAAAGGTTATGTCATCTAAAATGCTAGAATCAAAGGTAGGATATTCTACTAGAAAGATT 1357

Qy   1084 ttagaggatat 1094
Db   1358 TGTAAAGATTT 1368
```

```
RESULT   5
AF270394
LOCUS       AF270394     3493 bp    DNA             BCT       01-AUG-2000
DEFINITION  Staphylococcus epidermidis strain SR1 clone step.4046a02 genomic
            sequence.
ACCESSION   AF270394
VERSION     AF270394.1  GI:9624308
KEYWORDS    .
SOURCE      Staphylococcus epidermidis.
  ORGANISM  Staphylococcus epidermidis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE   1  (bases 1 to 3493)
  AUTHORS   Kimmerly,W.J., Taylor,J.David, Nelsen,A.J., Godlevski,M.M.,
            Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
            Listenbee,S., Ashanti,C., Altshuller,G., Mamo,L., Shepherd,N.S.,
            Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
            Furdon,P.J.
  TITLE     Transposon-mediated sequencing of the Staphylococcus epidermidis
            genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3493)
  AUTHORS   Taylor,J.David, Kimmerly,W.J., Nelsen,A.J., Godlevski,M.M.,
            Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
            Listenbee,S., Ashanti,C., Altshuller,G., Mamo,L., Shepherd,N.S.,
```

```
            Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
            Furdon,P.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2000) Departments of Genomic Sciences and
            Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
            Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES             Location/Qualifiers
     source          1. .3493
                     /organism="Staphylococcus epidermidis"
                     /strain="SR1"
                     /db_xref="taxon:1282"
                     /clone="step.4046a02"
BASE COUNT     1172 a    534 c    547 g   1240 t
ORIGIN

  Query Match             11.8%;  Score 135.4;  DB 2;  Length 3493;
  Best Local Similarity   46.7%;  Pred. No. 4.3e-22;
  Matches  510;  Conservative    0;  Mismatches  566;  Indels   15;  Gaps    2;

Qy       10 aataaaagagtattaatttgactgcaaattacggaaatggacatgtgcaggtagccaaa 69
Db      396 AATAAAAAGATATTGATTATTACTGGATCATTTGGTAACGGTCATATGCAAGTCACGCAA 455

Qy       70 acactttatgaacaatgtgtacggctcggctttcagcatgtaacagt------ttctaat 123
Db      456 AGTATTGTCAACCAATTGAATGAGATGAATCTCAATCATTTATCAGTCATTCAACATGAT 515

Qy      124 ttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaaagc 183
Db      516 TTGTTTATGGAAGCTCATCCAATTATGACTTCTATATGTAAGAAATGGTATATCAATAGC 575

Qy      184 ttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctataat 243
Db      576 TTTAAATATTTTAGAAATACATATAAACGATTTTACTATAGTCGCCCTAATGAGCTCGAT 635

Qy      244 aaacgtaaattcaatatttactttaaaatgggtaataaaagattgggcgaacttgtcgat 303
Db      636 AAATGTTTTTATAAATATTA---------TGGATTAAATAAACTCATCAACTTACTTATT 686

Qy      304 gaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatacaga 363
Db      687 AAAGAAAAGCCTGATCTCATATTATTAACATTTCCAACACCTGTGATGTCAGTGTTGACC 746

Qy      364 cgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcataaa 423
Db      747 GAACAATTTAATATAAATATCCCTATTGCGACAGTTATGACAGATTATCGCATGCATAAA 806

Qy      424 atttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaaaaa 483
Db      807 AATTGGATTACACCATATTCACAAAGATATTATGTAGCAACAAAAGATACTAAAGATGAT 866

Qy      484 ctgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcaggccg 543
Db      867 TTCATTGAAGCTGGTGTTCCTGCTTCATATATTAAAGTGACGGGCATTCCTATTGCTGAT 926

Qy      544 caatttgaagaatccatgcctgttggccgatatataaaaagtacaatctttcaccaaac 603
Db      927 AAATTTGAAGAATCTATTGATAAAGAAGAATGGTTATCGCAACAACATTTAGACCCTTCA 986

Qy      604 aaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagagctg 663
Db      987 AAACCTACTATATTAATGTCAGCAGGTGCATTTGGTGTTTCAAAAGGCTTTGACTATATG 1046

Qy      664 tgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaatacg 723
Db     1047 ATTAATAATATTTTAGAAAAAAGTCCAAATTCGCAAGTGGTCATGATTTGTGGACGTAGT 1106

Qy      724 gctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagttctg 783
Db     1107 AAGGAACTTAAACGTTCATTAAAAGCTAAGTTCAAAGATAATCCAAGTGTAATAATATTA 1166

Qy      784 ggctatgtggagcgcattgatgagctatttcggatcacagattgcatgattaccaagccc 843
Db     1167 GGATATACAAATCACATGAATGAGTGGATGGCATCAAGCCAACTAATGATTACAAAACCT 1226

Qy      844 ggcggcattactttgacagaagccacagccattggagtgcctgtcattctgtacaaaccc 903
Db     1227 GGTGGTATCACAATTTCCGAAGGACTTAGTCGTTGTATTCCTATGATTTTTTAAACCCT 1286

Qy      904 gtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcggagctgccatcgtt 963
Db     1287 GCACCCGGTCAAGAACTTGAAAATGCATATTACTTTGAAAGTAAAGGATTTGGAAAAATA 1346

Qy      964 gtgaaccgtcatgaagagattctcgagtcagtcacttcccttcttgcagatgaagatacc 1023
Db     1347 GCAGATACTCCAAATGAGGCAATTGATATTGTTTCTGACTTAACAAATAACGAAGAGACT 1406

Qy     1024 ttgcatcgcatgaagaaaaacattaaggaccttcatttagcaaactcctctgaagtgatt 1083
Db     1407 TTAAAGGTTATGTCATCTAAAATGCTAGAATCAAAGGTAGGATATTCTACTAGAAAGATT 1466

Qy     1084 ttagaggatat 1094
Db     1467 TGTAAAGATTT 1477


RESULT   6
SAY14370
LOCUS       SAY14370      7791 bp    DNA             BCT       24-JUN-1998
DEFINITION  Staphylococcus aureus RF3, murE, ypfP genes.
ACCESSION   Y14370
VERSION     Y14370.1  GI:3256221
KEYWORDS    murE gene; peptide chain release factor 3; RF3 gene;
            UDP-N-acetylmuramyl-tripeptide synthetase; ypfP gene.
SOURCE      Staphylococcus aureus.
  ORGANISM  Staphylococcus aureus
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE   1  (bases 1 to 7791)
  AUTHORS   Ludovice,A.M., Wu,S. and de Lencastre,H.
  TITLE     Molecular cloning and DNA sequencing of the Staphylococcus aureus
            UDP-N-acetylmuramyl tripeptide synthetase (murE) gene, essential
            for the optimal expression of methicillin resistance
  JOURNAL   Microb. Drug Res. 4, 85-90 (1998)
REFERENCE   2  (bases 1 to 7791)
  AUTHORS   Ludovice,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1997) A.M. Ludovice, Instituto de Tecnologia
            Qiomica e Biologica, Universidade Nova de Lisbon, R. da Quinta
            Grande 6, Apartado 127, 2780 Oeiras, PORTUGAL
  REMARK    Revised by author 16-JAN-98
COMMENT     Related sequence: L77246.
FEATURES             Location/Qualifiers
     source          1. .7791
                     /organism="Staphylococcus aureus"
                     /strain="COL"
                     /db_xref="taxon:1280"
     gene            complement(1435. .3114)
                     /gene="RF3"
     CDS             complement(1435. .3000)
                     /gene="RF3"
                     /codon_start=1
                     /transl_table=11
                     /product="peptide chain release factor 3"
                     /protein_id="CAA74739.1"
                     /db_xref="GI:3256222"
                     /db_xref="SPTREMBL:O86490"
                     /translation="MNLKQEVESRKTFAIISHPDAGKTTLTEKLLYFSGAIREAGTVK
                     GKKLVNLRQVTWMKVEQERGISVTSSVMQFDYDDYEINILDTPGHEDFSEDTYRTLMA
                     VDSAVMVIDCAKGVEPPTLKLFKVCKMRGIPIFTFINKLDRVGKEPFELLDEIEETLN
                     IETYPMNWPIGMGQSFFGIIDRKSKTIEPFRDEENILHLNDDFELEEDHAITNDSDFE
                     QAIEELMLVEEAGEAFDNDALLSGDLTPVFFGSALANFGVQNFLNAYVDFAPMPNARQ
                     TKENVEVSPFDDSFSGFIFKIQANMDPKHRDRIAFMRVVSGAFERVWMLLCNVLIKSK
                     RSHVQRHLWQTIKKLVNHAVAGDIIGLYDTGNYQIGDTLVGGKQTYSFQDLPQFTPEI
                     FMKVSAKNVMKQKHFHKGIEQLVQEGAIQYYKTLHTNQIILGAVGQLQFEVFEHRMKN
```

```
                EYNVDVVMEPVGRKIARVDIENEDQITDKMNTSRSILVKDRYDDLVFLFENEFATRWF
                EEKFPEIKLYSLL"
     RBS        complement(3009. .3013)
                /gene="RF3"
     -10_signal complement(3085. .3091)
                /gene="RF3"
     -35_signal complement(3109. .3114)
                /gene="RF3"
     gene       complement(3244. .4784)
                /gene="murE"
     CDS        complement(3244. .4722)
                /gene="murE"
                /codon_start=1
                /transl_table=11
                /product="UDP-N-acetylmuramyl-tripeptide synthetase"
                /protein_id="CAA74740.1"
                /db_xref="GI:3256223"
                /db_xref="SPTREMBL:O86491"
                /translation="MQVRCLKKVKVKRVLGSLEQQIDDITTDSRTAREGSIFVASVGY
                TVDSHKFCQNVADQGCKLVVVNKEQSLPANVTQVVVPDTLRVASILAHTLYDYPSHQL
                VTFGVTGTNGKTSIATMIHLIQRKLQKNSAYLGTNGFQINETKTKGANTTPETVSLTK
                KIKEAVDAGAESMTLEVSSHGLVLGRLRSVEFDVAIFSNLTQDHLDFHGTMEAYGHAK
                SLLFSQLGEDLSKEKYVVLNNDDSFSEYLRTVTPYEVFSYGIDEEAQFMAKNIQESLQ
                GVSFDFVTPFGTYPVKSPYVGKFNISNIMAAMIAVWSKGTSLETIIKAVENLEPVEGR
                LEVLDPSLPIDLIIDYAHTADGMNKLIDAVQPFVKQKLIFLVGMAGERDLTKTPEMGR
                VACRADYVIFTPDNPANDDPKMLTAELAKGATHQNYIEFDDRAEGIKHAIDIAEPGDT
                VVLASKGREPYQIMPGHIKVPHRDDLIGLEAAYKKFGGGPVD"
     RBS        complement(4736. .4740)
                /gene="murE"
     -10_signal complement(4758. .4763)
                /gene="murE"
     -35_signal complement(4779. .4784)
                /gene="murE"
     gene       5086. .6335
                /gene="ypfP"
     -35_signal 5086. .5092
                /gene="ypfP"
     -10_signal 5107. .5112
                /gene="ypfP"
     RBS        5145. .5150
                /gene="ypfP"
     CDS        5160. .6335
                /gene="ypfP"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA74741.1"
                /db_xref="GI:3256224"
                /db_xref="SPTREMBL:O86492"
                /translation="MVTQNKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHD
                LFMEAHPILTSICKKWYINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIK
                EKPDLILLTFPTPVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQ
                DFIDVGIDPSTVKVTGIPIDNKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGF
                DTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQL
                MITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVAS
                LTNGNEQLTNMISTMEQDKIKYATQTICRDLLDLIGHSSQPQEIYGKVPLYARFFVK"
BASE COUNT     2517 a   1452 c   1130 g   2668 t     24 others
ORIGIN
```

```
  Query Match            10.5%;  Score 120.4;  DB 3;  Length 7791;
  Best Local Similarity  48.3%;  Pred. No. 1.9e-18;
  Matches  473;  Conservative    0;  Mismatches  486;  Indels   21;  Gaps    4;

Qy     10 aataaaagagtattaattttgactgcaaattacggaaatggacatgtgcaggtagccaaa 69
          |||||||   |||| ||| | ||||     | ||| || || ||| |||| ||  |  |
Db   5172 AATAAAAAGATATTGATTATTACTGGCTCATTCGGTAACGGTCATATGCAAGTTACACAG 5231

Qy     70 acactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct------aat 123
          |   |   | | |||  |       |      |  | ||| |||  ||   |       ||
Db   5232 AGTATCGTTAATCAACTTAATGATATGAATCTAGACCATTTAAGCGTCATTGAGCACGAT 5291

Qy    124 ttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaaagc 183
          || |     ||  |  |||| ||| |  |      |   | || |    ||| | || |||

Db   5292 TTATTTATGGAAGCTCATCCAATTTTGACTTCTATTTGTAAAAAATGGTATATCAATAGC 5351

Qy    184 ttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctataat 243
          ||   |      | | |  | ||      ||||||||         |||    |    ||
Db   5352 TTTAAATATTTTAGAAATATGTACAAAGGGTTTTATT---------ACAGCCGCCCAGAT 5402

Qy    244 aaacgtaaattcaatatttacttttaaaatgggtaataaaagattgggcgaacttgtcgat 303
          ||||   |      | |||||    |    ||   ||| |  ||      |  | |
Db   5403 AAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTTATTGATA 5462

Qy    304 gaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatacaga 363
          || |  |||| ||| | ||  |    || ||||| |     ||  || ||| |   |
Db   5463 AAAGAAAAGCCAGATTTAATATTATTAACGTTTCCTACACCAGTTATGTCGGTACTAACT 5522

Qy    364 cgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcataaa 423
            | |  |    |     |||||       |  || ||||| || | | |  | ||||||
Db   5523 GAGCAATTTAACATTAATATTCCAGTTGCTACAGTGATGACAGACTATCGCTTACATAAA 5582

Qy    424 atttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaaaaa 483
          |  ||| ||       |       | ||||||||||||| ||| |  |   |||  || |
Db   5583 AACTGGATTACGCCGTATTCAACAAGATATTATGTGGCAACAAAAGAAACGAAACAAGAC 5642

Qy    484 ctgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcaggccg 543
          |  | ||  | || | | ||||    |  || ||| | ||||| ||||| ||
Db   5643 TTCATAGACGTAGGTATTGATCCTTCAACAGTTAAAGTGACAGGTATTCCTATTGATAAC 5702

Qy    544 caatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcaccaaac 603
          ||||||||    | ||   |      | |   |  | | |  |||| |    ||| |
Db   5703 AAATTTGAAACGCCTATTAATCAAAAGCAGTGGTTAATAGACAACAACTTAGATCCAGAT 5762

Qy    604 aaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagagctg 663
          ||  ||    || | ||    || |||||    |||||||  ||      |  |   | ||
Db   5763 AAGCAAACTATTTTAATGTCAGCTGGTGCATTTGGTGTATCTAAAGGTTTTGACACGATG 5822

Qy    664 ---tgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720
             || |   | |  ||   ||   | |  |||||||||  |  | || || || | |
Db   5823 ATTACTGATATATTAGCGAAAAGTGCAAATGCACAAGTAGTTATGATTTGTGGTAAGAGC 5882

Qy    721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
          |  |   ||||    ||||| |  ||          ||   |     | || |   ||
Db   5883 AAAGAGCTAAAGCGTTCTTTAACAGCTAAGTTTAAATTAACGAGAATGTATTTGA---TT 5939

Qy    781 ctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgattaccaag 840
          || || |||    | | |||  ||||      |       |   |    |||||||| ||
Db   5940 CTAGGTTATACCAAACACATGAATGAATGGATGGCATCAAGTCAACTTATGATTACGAAA 5999

Qy    841 cccggcggcattactttgacagaagccacagccattggagtgcctgtcattctgtacaaa 900
          || || ||  || ||  | || ||||     |||  | |  | ||  | |||  |    ||
Db   6000 CCTGGTGGTATCACAATAACTGAAGGTTTCGCCCGTTGTATTCCAATGATTTTCCTAAAT 6059

Qy    901 cccgtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcggagctgccatc 960
          || |  |||| || ||    |||||||    | |||||||||    ||   ||  |
Db   6060 CCTGCACCTGGTCAAGAGCTTGAAAATGCCTTTTACTTTGAAGAAAAAGGTTTTGGTAAA 6119

Qy    961 gttgtgaaccgtcatgaaga 980
          | |  |   ||  ||||
Db   6120 ATCGCTGATACTCCAGAAGA 6139
```

```
RESULT   7
AX016297
LOCUS       AX016297      975 bp    DNA             PAT       07-SEP-2000
DEFINITION  Sequence 2 from Patent WO9949052.
ACCESSION   AX016297
VERSION     AX016297.1  GI:10041860
KEYWORDS    .
SOURCE      Staphylococcus aureus.
  ORGANISM  Staphylococcus aureus
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE   1  (bases 1 to 975)
  AUTHORS   Zaehringer,U., Heinz,E., Jorasch,P. and Wolter,F.P.
  TITLE     Processive glycosyltransferase
```

```
JOURNAL    Patent: WO 9949052-A 2 30-SEP-1999;
           FORSCHUNGSZENTRUM BORSTEL (DE); ZAEHRINGER ULRICH (DE); HEINZ ERNST
           (DE); JORASCH PETRA (DE); WOLTER FRANK P (DE); GVS GES FUER ERWERB
           UND VERWER (DE)
FEATURES             Location/Qualifiers
     source          1. .975
                     /organism="Staphylococcus aureus"
                     /db_xref="taxon:1280"
BASE COUNT      350 a    152 c    169 g    304 t
ORIGIN

  Query Match             10.3%;  Score 118;  DB 9;  Length 975;
  Best Local Similarity   48.5%;  Pred. No. 5.6e-18;
  Matches  458;  Conservative    0;  Mismatches  465;  Indels   21;  Gaps    4;

Qy      10 aataaaagagtattaattttgactgcaaattacggaaatggacatgtgcaggtagccaaa 69
Db      13 AATAAAAAGATATTGATTATTACTGGCTCATTCGGTAACGGTCATATGCAAGTTACACAG 72

Qy      70 acactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct------aat 123
Db      73 AGTATCGTTAATCAACTTAATGATATGAATCTAGACCATTTAAGCGTCATTGAGCACGAT 132

Qy     124 ttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaaagc 183
Db     133 TTATTTATGGAAGCTCATCCAATTTTGACTTCTATTTGTAAAAAATGGTATATCAATAGC 192

Qy     184 ttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctataat 243
Db     193 TTTAAATATTTTAGAAATATGTACAAAGGGTTTTATT---------ACAGCCGCCCAGAT 243

Qy     244 aaacgtaaattcaatatttactttaaaatgggtaataaaagattgggcgaacttgtcgat 303
Db     244 AAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTTATTGATA 303

Qy     304 gaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatacaga 363
Db     304 AAAGAAAAGCCAGATTTAATATTATTAACGTTTCCTACACCAGTTATGTCGGTACTAACT 363

Qy     364 cgccgaactggaagagtcattcctaccttcaacgttatgactgattttgtcttcataaa 423
Db     364 GAGCAATTTAACATTAATATTCCAGTTGCTACAGTGATGACAGACTATCGCTTACATAAA 423

Qy     424 atttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaaaaa 483
Db     424 AACTGGATTACGCCGTATTCAACAAGATATTATGTGGCAACAAAAGAAACGAAACAAGAC 483

Qy     484 ctgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcaggccg 543
Db     484 TTCATAGACGTAGGTATTGATCCTTCAACAGTTAAAGTGACAGGTATTCCTATTGATAAC 543

Qy     544 caatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcaccaaac 603
Db     544 AAATTTGAAACGCCTATTAATCAAAAGCAGTGGTTAATAGACAACAACTTAGATCCAGAT 603

Qy     604 aaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagagctg 663
Db     604 AAGCAAACTATTTTAATGTCAGCTGGTGCATTTGGTGTATCTAAAGGTTTTGACACGATG 663

Qy     664 tgc---gaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720
Db     664 ATTACTGATATATTAGCGAAAAGTGCAAATGCACAAGTAGTTATGATTTGTGGTAAGAGC 723

Qy     721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
Db     724 AAAGAGCTAAAGCGTTCTTTAACAGCTAAGTTTAAATTAACGAGAATGTATTTGA---TT 780

Qy     781 ctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgattaccaag 840
Db     781 CTAGGTTATACCAAACACATGAATGAATGGATGGCATCAAGTCAACTTATGATTACGAAA 840

Qy     841 cccggcggcattactttgacagaagccacagccattggagtgcctgtcattctgtacaaa 900
Db     841 CCTGGTGGTATCACAATAACTGAAGGTTTCGCCCGTTGTATTCCAATGATTTTCCTAAAT 900

Qy     901 cccgtgcctggccaggaaaaagaaaatgcaaacttctttgaaga 944
Db     901 CCTGCACCTGGTCAAGAGCTTGAAAATGCCTTTTACTTTGAAGA 944

RESULT 8
AP001513
LOCUS       AP001513   292550 bp    DNA             BCT       10-JAN-2001
DEFINITION  Bacillus halodurans genomic DNA, section 7/14.
ACCESSION   AP001513 BA000004
VERSION     AP001513.1  GI:10174345
KEYWORDS    .
SOURCE      Bacillus halodurans DNA.
  ORGANISM  Bacillus halodurans
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (sites)
  AUTHORS   Takami,H., Nakasone,K., Hirama,C., Takaki,Y., Masui,N., Fuji,F.,
            Nakamura,Y. and Inoue,A.
  TITLE     An improved physical and genetic map of the genome of alkaliphilic
            Bacillus sp. C-125
  JOURNAL   Extremophiles 3 (1), 21-28 (1999)
   MEDLINE  99184645
REFERENCE   2  (sites)
  AUTHORS   Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y.,
            Masui,N., Fuji,F., Takaki,Y., Inoue,A. and Horikoshi,K.
  TITLE     Sequencing of three lambda clones from the genome of alkaliphilic
            Bacillus sp. strain C-125
  JOURNAL   Extremophiles 3 (1), 29-34 (1999)
   MEDLINE  99184646
REFERENCE   3  (sites)
  AUTHORS   Takami,H., Takaki,Y., Nakasone,K., Hirama,C., Inoue,A. and
            Horikoshi,K.
  TITLE     Sequence analysis of a 32-kb region including the major ribosomal
            protein gene clusters from alkaliphilic Bacillus sp. strain C-125
  JOURNAL   Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)
   MEDLINE  99209008
REFERENCE   4  (sites)
  AUTHORS   Takami,H., Masui,N., Nakasone,K. and Horikoshi,K.
  TITLE     Replication origin region of the chromosome of alkaliphilic
            Bacillus halodurans C-125
  JOURNAL   Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)
   MEDLINE  99356711
REFERENCE   5  (sites)
  AUTHORS   Takami,H., Takaki,Y., Nakasone,K., Sakiyama,T., Maeno,G.,
            Sasaki,R., Hirama,C., Fuji,F. and Masui,N.
  TITLE     Genetic analysis of the chromosome of alkaliphilic Bacillus
            halodurans C-125
  JOURNAL   Extremophiles 3 (3), 227-233 (1999)
   MEDLINE  99411980
REFERENCE   6  (sites)
  AUTHORS   Takami,H.
  TITLE     Genome analysis of facultatively alkalihilic Bacillus halodurans
            C-125
  JOURNAL   (in) Extremophiles in deep-sea environments (Ed.);
            . HORIKOSHI, K. TSUJII;
            : 249-284; Springer-Verlag (1999)
REFERENCE   7  (sites)
  AUTHORS   Takami,H. and Horikoshi,K.
  TITLE     Reidentification of facultatively alkaliphilic Bacillus sp. C-125
            to Bacillus halodurans
  JOURNAL   Biosci. Biotechnol. Biochem. 63, 943-945 (1999)
REFERENCE   8  (sites)
  AUTHORS   Takami,H., Nakasone,K., Takaki,Y., Maeno,G., Sasaki,R., Masui,N.,
            Fuji,F., Hirama,C., Nakamura,Y., Ogasawara,N., Kuhara,S. and
            Horikoshi,K.
  TITLE     Complete genome sequence of the alkaliphilic bacterium Bacillus
            halodurans and genomic sequence comparison with Bacillus subtilis
  JOURNAL   Nucleic Acids Res. 28, 4317-4331 (2000)
REFERENCE   9  (sites)
```

```
  AUTHORS   Takami,H. and Horikoshi,K.
  TITLE     Analysis of the genome of an alkaliphilic Bacillus strain from an
            industrial point of view
  JOURNAL   Extremophiles 4 (2), 99-108 (2000)
  MEDLINE   20263314
REFERENCE   10 (sites)
  AUTHORS   Nakasone,K., Masui,N., Takaki,Y., Sasaki,R., Maeno,G., Sakiyama,T.,
            Hirama,C., Fuji,F. and Takami,H.
  TITLE     Characterization and comparative study of the rrn operons of
            alkaliphilic Bacillus halodurans C-125
  JOURNAL   Extremophiles 4 (4), 209-214 (2000)
  MEDLINE   20426005
REFERENCE   11 (bases 1 to 292550)
  AUTHORS   Takami,H. and Takaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and
            Technology Center, Deep-sea Microorganisms Research Group; 2-15
            Natsushima, Yokosuka, Kanagawa 237-0061, Japan
            (E-mail:takamih@jamstec.go.jp,
            URL:http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/FResearch.html,
            Tel:81-468-67-3895, Fax:81-468-66-6364)
FEATURES             Location/Qualifiers
     source          1. .292550
                     /organism="Bacillus halodurans"
                     /db_xref="taxon:86665"
                     /note="alkaliphile"
     gene            225. .4817
                     /gene="gltA"
     CDS             225. .4817
                     /gene="gltA"
                     /note="BH1728"
                     /codon_start=1
                     /transl_table=11
                     /product="glutamate synthase (large subunit)"
                     /protein_id="BAB05447.1"
                     /db_xref="GI:10174346"
                     /translation="MQNYGYPKKQGLYDPQFEHDNCGIGFLAHMKGQKSHKIVEDALH
                     ILCNLEHRGGQGDEVNTGDGAGILLQLPHRFFQKVTVSEGISLPEEGKYGVGMLFLPQ
                     DESLRRETERKLEAIVNEEGQTFLGWRTVPTDDSMLGNAAKAAMPYIRQLFIQKSENV
                     KAGLDFERKLYVIRKRAENEIGPLVESGSFYFASFSSRTIVYKGMLTTEQVDQFYLDV
                     VDPDFETALALVHSRFSTNTFPSWERAHPNRYMIHNGEINTVKGNVNWMHAREARFES
                     EVFGTDLNKIHPVIDQEGSDSSMFDNTIEFLSLSGRSMAHAAMMMIPEPWQNDEHMSK
                     EKKAFYQFHSTLMEPWDGPTAIVFTDGTQIGACLDRNGLRPSRYYVTKDDYIIMSSEV
                     GVLDIDPENVLYKERLHPGHMLLVDLEEGRIIPDDEIKHQIASEHPYEEWVKENLVQL
                     EDVLETSHVEDTDFGKLRERQLAFGYTYEELTKMIQPMVKEGEDPVGSMGYDSPLAVL
                     SKRPQLLFNYFKQLFAQVTNPPIDAIREKVVTAVGTTIGAERNLLNPEPLSCRHIYLP
                     SPILNNEELAKLKHNKMDGFTSTTLSILFDAKSGAKGLEAALENLFSEADQAIEAGKT
                     LLILSDRKMDETNAAIPSLLAVSGLHHHLIRKGTRTKVSILLESGEPREVHHHCVLLG
                     YGAEAINPYLVFDSIDALLRDGLLDNISYVEAVENYVKAAVKGIMKVLSKMGISTIQS
                     YRGAQIFEAVGVDESVINQYFTWTTTRIGGVTLKEIAEEVRLRHERAFSEQEGVDKAL
                     DPGDDLQWRRNGDPHQYNPHTIHMLQHACRKNDYDLYKKYASAINEQEKEQTTLRGLL
                     KFTTDRKPISLEEVEPVEEIVKRFKTGAMSFGSISKEAHETLAIAMNRLGGKSNTGEG
                     GEDPDRFTPDANGDLRRSAIKQVASGRFGVTSHYLVNADEIQIKVAQGAKPGEGGQLP
                     GNKVYPWVAEVRGSTPGVGLISPPPHHDIYSIEDLAELIHDLKNANPQARISVKLVAG
                     TGVGTIAAGVAKGRADHVVISGYDGGTGAAARTSIKHAGLPWEIGLAETHQTLVLNNL
                     RDRISIETDGKLMTGRDVVIAALLGAEEYAFSTAPLVVLGCIIMRVCHLDTCPVGIAT
                     QNPELRKKYMGKPEHVVNYMTFVAQEIREIMAELGFRTVDEMIGRTDVLEQNRNIDNW
                     KAVHVDLSSLLYQPKPGQQEKNYKTIEQDHQLEQSLDMTKLLDICRPALERKEKVQAA
                     VSIKNIDRVAGTIVGSEISKRYGAAGLPEDTIELTFKGSAGQSFGAFIPRGMTLHLEG
                     DANDYTAKGLSGGKIIVRPPAGSTFKAEENVIIGNTSFYGATSGEAFIRGQAGERFCV
                     RNSGVNVVIEGMGDHGLEYMTGGRVINLGKVGKNFGAGMSGGVGYIYDYDGTFEKKCN
                     KEMILLESLSDEGDVVFIKSMLERHLAYTDSDLAKSILADWNQARKRFIKVIPKDYKR
                     MLEAIERVKREGLSDTEAVMVAFDENKNDKSRVGGK"
     gene            4912. .6399
                     /gene="gltB"
     CDS             4912. .6399
                     /gene="gltB"
                     /note="BH1729"
                     /codon_start=1
                     /transl_table=11
                     /product="glutamate synthase (small subunit)"
                     /protein_id="BAB05448.1"
                     /db_xref="GI:10174347"
                     /translation="MGKPTGFMEYKRETPPKKDPFERSKNWQEFTLLMPESALRQQGA
                     RCMDCGVPFCQTGTSMDGSGEIGCPVYNLIPEWNDLVYRGHWKEALERLHKTNNFPEF
                     TGRVCPAPCEGSCTVAISDDAVAIKSMEFHIVERGFQEGWIVPNPPRSRTGKKVAVVG
                     SGPAGLAAAAQLNKAGHWVTVFEREDRIGGLLTYGIPDVKLPYHVVERRINLLKEEGI
                     EFRTNVEIGKDVSAKQLKDDFDAVVLCTGATKPRDVNVKGRELKGIHFAMDFLTANTK
                     SLLNSNHQDGNYISAKGKHVVVIGGGDTGADCITTSVRHGAKSITQFDINREKSNERP
                     DNNPWPLFPIVHTVEDGHKEAAAVYGKDPRAYQVMTTKFVGDAEGHVKEIHTISVETT
                     VHPDGSKTRVEIPGTEKVWPADLVFLAVGFTGPEEQLVRELELETDTRSNIKAEYGKY
                     ETSLEGVFAAGDNRRGQSLVVWAIHEGREAARECDKYLMGASNLP"
     gene            6594. .7232
                     /gene="BH1730"
     CDS             6594. .7232
                     /gene="BH1730"
                     /note="BH1730
                     unknown conserved protein in B. subtilis"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB05449.1"
                     /db_xref="GI:10174348"
                     /translation="MVKWMKRGVLVSTALITLGFASPTDVVRGTENSSLPEKSENTDN
                     GKPIIYRVDADSYGLYDSVLKTPNHYVHLADETTARLFADGIIEEAYEQSLKKFGPVI
                     ARKIGSMYEEKILPEFKHAIVDAVAGMPVDDLNDLKVSDDPASGQGEKIVHLYHQKTG
                     NDLLRLHVRRDQPPKQGYWFNFHYHSYEDQFQSHRELGTVFWGKNMPPYWQA"
     gene            7873. .8241
                     /gene="BH1731"
     CDS             7873. .8241
                     /gene="BH1731"
                     /note="BH1731
                     unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB05450.1"
                     /db_xref="GI:10174349"
                     /translation="MVKQWSMILFAFVLFIGVWALAVIKLNSESSAPQEMLSDEVVER
                     EDNLTYQFGRDDSQAINLQESQYEESEVIEKHEKNNEKVERSIKRDDKWIDLRKYDLS
                     VISTPSGVPIDDLLDLLEIQ"
     gene            8329. .8580
                     /gene="BH1732"
     CDS             8329. .8580
                     /gene="BH1732"
                     /note="BH1732
                     unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB05451.1"
                     /db_xref="GI:10174350"
                     /translation="MEMKQWLRTKSPLLKWVLRLLFAYVCVIVATLALAIISIFITFT
                     FVIIDTISDSTHLGDFAEAYFVPVSEYLWDMFLWLVPGL"
     gene            8601. .8780
                     /gene="BH1733"
     CDS             8601. .8780
                     /gene="BH1733"
                     /note="BH1733
                     unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB05452.1"
                     /db_xref="GI:10174351"
                     /translation="MTEEIIIVLIVFGFFLMTGVCGGLGVFGMLTGRRKLAKWGLATG
                     FAFIAIYVVFMFMRL"
     gene            8796. .8954
                     /gene="BH1734"
     CDS             8796. .8954
                     /gene="BH1734"

  Query Match           6.1%;  Score 69.6;  DB 2;  Length 292550;
  Best Local Similarity  53.3%;  Pred. No. 4.1e-06;
  Matches  147;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy    823 gattgcatgattaccaagcccggcggcattactttgacagaagccacagccattggagtg 882
          |||   ||  |  | ||| |||| ||  |  ||  ||  | | |||  |   ||    ||    |
Db 253846 GATGCGATTTTGTCTAAGGCCGGGGGGTTAACGATGGCTGAATCATTAGTGTGTGAGACG 253905
```

```
Qy    883 cctgtcattctgtacaaacccgtgcctggccaggaaaaagaaaatgcaaacttctttgaa 942
          ||  | ||  |  || | || ||||||| || ||  || | || ||||| || ||
Db 253906 CCAATTATCATTCACCAGCCAGTGCCTGGTCATGAGGAACACAACGCAAAATTTTTAATT 253965

Qy    943 gaccgcggagctgccatcgttgtgaaccgtcatgaagagattctcgagtcagtcacttcc 1002
          |||   || || ||  |    || ||  |      | || ||||      || |||
Db 253966 GACGCTGGCGCAGCGCTAAGGGTAAAGGGAAGCAAGGAAATTCCTACAACAATCAAGCGA 254025

Qy   1003 cttcttgcagatgaagataccttgcatcgcatgaagaaaaacattaaggaccttcattta 1062
           |||||   || ||||    |||   ||  ||||   ||||    |   | ||  |   |
Db 254026 GTTCTTTACGAAGAAGCGTGCTTTGGTCCGATGATCGAAAATGCGAGAAAACTAAAAAAA 254085

Qy   1063 gcaaactcctctgaagtgattttagaggatatcctg 1098
          ||||  |  |  | | |||| | ||| | || |||
Db 254086 CCAAATGCTGCCAATGAGATTGTGGAGCAGATGCTG 254121
```

```
RESULT   9
AE001958
LOCUS       AE001958    13408 bp    DNA             BCT       22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 95 of 229 of the complete
            chromosome 1.
ACCESSION   AE001958 AE000513
VERSION     AE001958.1  GI:6458805
KEYWORDS    .
SOURCE      Deinococcus radiodurans.
  ORGANISM  Deinococcus radiodurans
            Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
REFERENCE   1  (bases 1 to 13408)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et,al.
  TITLE     Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
  JOURNAL   Science 286 (5444), 1571-1577 (1999)
  MEDLINE   20036896
REFERENCE   2  (bases 1 to 13408)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .13408
                     /organism="Deinococcus radiodurans"
                     /strain="R1"
                     /db_xref="taxon:1299"
                     /chromosome="1"
     gene            82. .447
                     /gene="DR1073"
     CDS             82. .447
                     /gene="DR1073"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF10658.1"
                     /db_xref="GI:6458817"
                     /translation="MTTLLLIVGLSVLLLGLMLLSRPKRPPQETQPLAPRRVRWLDPA
                     VQAVFDRSSGKTTPVRRGQRVVWEPRVLLKLRGDRAEMERLILQQTDLFPRASREQLL
                     KIIYDTRCGIGTCRSGSDA"
     gene            431. .1186
                     /gene="DR1074"
     CDS             431. .1186
                     /gene="DR1074"
                     /note="similar to SP:P32205 PID:397538 percent identity:
                     72.14; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="(3R)-hydroxymyristoyl-acyl carrier protein
                     dehydratase"
                     /protein_id="AAF10647.1"
                     /db_xref="GI:6458806"
                     /translation="MVQTLRMPPMRWLPSVLLLLTGLLLAALALGVFAVLGGSGPLWL
                     RSLGSVTLGGPGPNGFARGLALMLLASTRPRAGDAAPSGTQIFRTRARTGARAAAGLR
                     SGLTMSLMTPLLIRDVLKALPHRYPFVLVDRVFSTENGEVHALKNVTINEPFFMGHFP
                     TEPVMPGVLITEALAQASMFCLHGQMEPGQIGYLAGIEGARFKRKVIPGDQLHLHAKL
                     EFLRRGLGKTTCRAEVDGEVAAEMQILFAVAKG"
     gene            1359. .2543
                     /gene="DR1075"
     CDS             1359. .2543
                     /gene="DR1075"
                     /note="similar to PID:1001350 PID:1001356 percent
                     identity: 57.57; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAF10648.1"
                     /db_xref="GI:6458807"
                     /translation="MFLSDPLPAGGDPVTRLTRYVTRELLPPLLAGTLLFTAILSFGY
                     FFISSQWLSGVPVSLIARWIGYQLPDTLVKVLPMAVVLMTVVAFGRMNTERELVAVQA
                     GGIGLGRVTRPAAVVGLLVTALSVWLSLWVAPRANVETRGLYWDALTGAGLSQLVGKT
                     VDLGQGLTLAMQGYDAGARQMQGVRVEKWDKDNFRQGMVVFADAGKFEGNQLSLTGYQ
                     VYRVDYAAAAALSGVADNDPAAFREAVQKVFPLVQVPEDRNAALNLDTGLSRKQTLAQ
                     YADAIGADAEGWPELITKLTAPGVPQSERDAARVNLNRKLALPFANLVLVLAALPFAL
                     RFGRTLGVSLGVALVIAVAYYLLFSVGLTLAGLLPGLPELGVWLANIVFALGGLWLLR
                     RA"
     gene            2486. .3721
                     /gene="DR1076"
     CDS             2486. .3721
                     /gene="DR1076"
                     /note="similar to SP:P54166 PID:1256630 GB:AL009126
                     percent identity: 48.75; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="cell wall synthesis protein, putative"
                     /protein_id="AAF10649.1"
                     /db_xref="GI:6458808"
                     /translation="MARQHRLRAGRPVAAEACMSGGETGRTLRALFMSVSLGAGHDQA
                     QQAVKQAFAERGVELLGAEHDSVEYLSTFERSFTVDLYEFELRYAPWLYRGFYWLTDQ
                     DQPWNIISRMFTWLGMGAFKDELRELRPEVVINSFWAPAAVCDTLRAQTGQRFLNCLI
                     VTDYRAHLHWARRETDLLMVASEETRRQMLERGVRPEQVEVTGIPISPAFREVLAADR
                     WALRAELFSEMSLRPGVPLLLLSGGGRGHYAAAADVLTELGNLGRAVQVLVPASRQGE
                     GTETIGGATVHHLGFRRDLPRLLAASDLVVGKAGGLTVAEATALGVPLVIYAPIPGQE
                     EHNADFLERHGAGLWARARHDVRPLVLRALDPAEHARLSAGARAVGIPDAADRVAGAI
                     LRRLGDGLGDELGDSRGPA"
     gene            3718. .4923
                     /gene="DR1077"
     CDS             3718. .4923
                     /gene="DR1077"
                     /note="similar to PID:595264 SP:P50352 percent identity:
                     52.00; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="chitooligosaccharide deacetylase"
                     /protein_id="AAF10650.1"
                     /db_xref="GI:6458809"
                     /translation="MKRGVRGLLLGAALYIGLPYLLVQVGNLGLVREGRRARREVALT
                     FDDGPDPQTTPAVLAALREADMHATFFVIAGKAQAHPDLIRQMLEEGHEVEAHADKHV
                     HAWIRTPWGAALDPLRAVRAVGAMTGRPVRFHRPPHGAYTLSTWLGQRLAGVRGAHWS
                     IEGCDWHPESIPDTVRERLAALLVPGAVIVLHDAGPGARVTVPLLPSLLADLKARGYR
                     SVTLAELGGAAPQDWPGLKRRGFLALDAVFDRLGHIHFAGGRADNLFRIARVPFPLEG
                     ARLADGTPIPHGAPALEFHVNNPILVDLGPRASVRQARREDFRVVARELQTRPEYADV
                     GYVFCLSAVSPLLGLLGFENHDLPAADARRLRRWANVLRRAYGNDPNAKAPRLSVLTR
                     EEFLALYGS"
     gene            4929. .5819
                     /gene="DR1078"
```

```
CDS             4929. .5819
                /gene="DR1078"
                /note="similar to GB:X79146 PID:487700 percent identity:
                46.92; identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="antibiotic biosynthesis protein LmbJ, putative"
                /protein_id="AAF10651.1"
                /db_xref="GI:6458810"
                /translation="MRAVGLLTLYPCGRGPAEGRGEGGFLGQMLYEWGSKGRLCFDSL
                TPYPPPVNYDDLAPIYDQQYDSYRDDLHFYAGLAERAGGRVLEIGAGTGRVTAFLTRR
                GAAVLGVEPSGEMIVGAQARAAREGLTLELVQATAQTFASDERFGLIIAPFNALMHLY
                TPAEQLAALQNIRAHLAPGGQFVFDLYVPHFGAMNTLRHEGETFHVPDGSRTDLFLLQ
                RHDAPRQVITTEYFADTTAPDGALRRAHHTLTQRYYTRFEMEWLLRCAGFEAPRVTGS
                FQGGPLVETSEVMVFQARGA"
gene            complement(5801. .6058)
                /gene="DR1079"
CDS             complement(5801. .6058)
                /gene="DR1079"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF10659.1"
                /db_xref="GI:6458818"
                /translation="MKLTEDQGAHITLDITRDELTVLAGGLLTALDLLTRADAAEDLR
                LHLEAEFHGRVGVSRASATALLTELVGIIQAGEDAAQAPLA"
gene            complement(6135. .6794)
                /gene="DR1080"
CDS             complement(6135. .6794)
                /gene="DR1080"
                /note="similar to SP:P37786 PID:454907 percent identity:
                59.86; identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="rfbJ protein/conserved hypothetical protein"
                /protein_id="AAF10652.1"
                /db_xref="GI:6458811"
                /translation="MSAVPTRQQPTVAVVIPAFNEEETVGDVVRAALTLTPEVVVASD
                GSADATAQVARTAGARVVELLENAGKGPALAAALQATDAEYVVMLDGDLQGLTREHLD
                VLLRPVLDGELDMTIGIFAGGGFASDFGNKMTPQLSGQRACRRDWLLGVPRLAAERWP
                EPAITSHLKATDARWAYVELPQVAQVLKETKRGFWSGVKARSKMYGQLLTFWVRKRKD
                D"
gene            complement(6852. .7553)
                /gene="DR1081"
CDS             complement(6852. .7553)
                /gene="DR1081"
                /note="similar to PID:868019 GB:U00096 SP:Q59431
                PID:1742672 PID:1742680 percent identity: 51.90;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="transcriptional regulator, TetR family"
                /protein_id="AAF10653.1"
                /db_xref="GI:6458812"
                /translation="MKVDRQEQDEARRTRIARAAFELFARTGLDSVSAQDIARAAFVS
                RTNLYRYFPSKTHMLLAHFERTVGETRAEALRRLSSGANPQVVWQLVTTRMADLGVRY
                RHLAGAVGQAVLGRSRPAEGTPSGPVPQELQTAQTLVSLVEPVLSAMRQQGKLRPDAD
                TRFLAALLVDACLMSLLHGGHRNQREVLGDWQDRFSLLMHGALAPNAPALPPDGQRPP
                ASSDEPPTPPLSESR"
gene            7805. .8380
                /gene="DR1082"
CDS             7805. .8380
                /gene="DR1082"
                /note="similar to GB:Z31376 SP:P28368 PID:216333
                PID:499384 PID:1762348 percent identity: 67.21; identified
                by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="light-repressed protein A, putative"
                /protein_id="AAF10654.1"
                /db_xref="GI:6458813"
                /translation="MQIYQLSGRNVEVTEPMREYVEEKLSRLDRYTDQITDARVTLTV
```

```
  Query Match             4.7%;  Score 53.8;  DB 1;  Length 13408;
  Best Local Similarity   62.0%;  Pred. No. 0.018;
  Matches   85;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy     819 cacagattgcatgattaccaagcccggcggcattactttgacagaagccacagccattgg 878
           | | ||   | || |   |||| | |||||| | ||  || | || || || ||  | ||
Db    3388 CTCCGACCTCGTGGTCGGCAAGGCGGGCGGCCTGACGGTGGCCGAGGCGACGGCTCTCGG 3447

Qy     879 agtgcctgtcattctgtacaaacccgtgcctggccaggaaaaagaaaatgcaaacttctt 938
           |||||  || |  | |||   ||  | ||||| |||||| || | || ||  ||||  |
Db    3448 CGTGCCGCTCGTCATCTACGCGCCGATTCCTGGTCAGGAAGAACACAACGCCGACTTTCT 3507

Qy     939 tgaagaccgcggagctg 955
           |||  || ||| || |
Db    3508 GGAACGCCACGGTGCCG 3524


RESULT  10
I66494/c
LOCUS       I66494       7218 bp    DNA             PAT       28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             4.3%;  Score 49;  DB 10;  Length 7218;
  Best Local Similarity   2.7%;  Pred. No. 0.24;
  Matches   10;  Conservative  216;  Mismatches  151;  Indels    0;  Gaps    0;

Qy     419 ataaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaagg 478
           | || | || ||| |: ::::: : ::: ::: :  : : :: :: :::  : : :::::
Db    1449 AGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1390

Qy     479 aaaaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatca 538
           ::::: :  ::::: :: :   :   ::: :: : ::::  : :::::    ::  :
Db    1389 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1330

Qy     539 ggccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcac 598
           ::  : ::   ::::::   : :   :  ::   :: : : :::::: : ::      :
Db    1329 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1270

Qy     599 caaacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaag 658
           ::: :::::: :     ::  : :: ::: :   : :: : :  :::::: : :::::
Db    1269 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1210

Qy     659 agctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaa 718
           ::  : : ::::     :  ::::: ::  ::: : ::: ::   : : : : :::::::
Db    1209 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1150

Qy     719 atacggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaag 778
           : : ::    :::::::     ::: : :   :::: :::::: :: :: :::  :::::
Db    1149 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1090

Qy     779 ttctgggctatgtggag 795
              :::  : : ::::
Db    1089 RRRRRRRRRRRRRRRR 1073


RESULT  11
```

```
E14811
LOCUS       E14811       1266 bp    DNA             PAT       28-JUL-1999
DEFINITION  cDNA encoding monogalactosyl diacyl glycerol synthase.
ACCESSION   E14811
VERSION     E14811.1  GI:5709494
KEYWORDS    JP 1998014579-A/2.
SOURCE      Cucumis sativus.
  ORGANISM  Cucumis sativus
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Cucurbitales; Cucurbitaceae; Cucumis.
REFERENCE   1  (bases 1 to 1266)
  AUTHORS   Ota,H., Shimojima,M., Iwamatsu,A. and Takamiya,K.
  TITLE     MONOGALACTOSYL DIACYLGLYCEROL SYNTHASE AND GENE CODING THE SAME
  JOURNAL   Patent: JP 1998014579-A 2 20-JAN-1998;
            KIRIN BREWERY CO LTD
COMMENT     OS   Cucumis sativus
            PN   JP 1998014579-A/2
            PD   20-JAN-1998
            PF   02-JUL-1996 JP 1996172337
            PI   OTA HIROYUKI, SHIMOJIMA MIE, IWAMATSU AKIHIKO, PI   TAKAMIYA
            KENICHIRO
            PC   C12N15/09,C07H21/04,C07K14/415,C07K19/00,C12N1/21,C12N9/10, PC
              (C12N1/21,
            PC   C12R1:19),(C12N9/10,C12R1:19);
            CC   strandedness: Double;
            CC   topology: Linear;
            CC   hypothetical: No;
            CC   anti-sense: No;
            FH   Key             Location/Qualifiers
            FH
            FT   source          1. .1266
            FT                   /organism='Cucumis sativus'.
FEATURES             Location/Qualifiers
     source          1. .1266
                     /organism="Cucumis sativus"
                     /db_xref="taxon:3659"
BASE COUNT      362 a    238 c    319 g    347 t
ORIGIN
```

```
  Query Match             4.2%;  Score 48;  DB 10;  Length 1266;
  Best Local Similarity   51.9%;  Pred. No. 0.35;
  Matches  108;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy     776 aagttctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgatta 835
           | ||   ||| | |||       || || ||     |  |       |||||||| || |
Db     899 AGGTGAAGGGGTTTGTCACAAAAATGGAGGAATGCATGGGAGCTTGTGATTGCATTATAA 958

Qy     836 ccaagcccggcggcattactttgacagaagccacagccattggagtgcctgtcattctgt 895
           | ||| | |||      ||  |  | |||||||     |  ||  | ||| | ||||||
Db     959 CAAAGGCGGGCCCTGGAACGATTGCTGAAGCCATGATAAGAGGTCTTCCTATAATTCTGA 1018

Qy     896 acaaacccgtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcggagctg 955
           |  |   | |  |||| ||||||   | |||||     |   | |  ||   ||||  ||
Db    1019 ATGACTACATTGCTGGACAGGAAGCTGGAAATGTGCCATATGTCGTCGAAAACGGATGTG 1078

Qy     956 ccatcgttgtgaaccgtcatgaagagat 983
           |   ||   ||   || | |||||||
Db    1079 GGAAGTTTTCAAAATCTCCTAAAGAGAT 1106
```

```
RESULT  12
CSU62622
LOCUS       CSU62622     2142 bp    mRNA            PLN       28-JAN-1997
DEFINITION  Cucumis sativus monogalactosyldiacylglycerol synthase mRNA,
            complete cds.
ACCESSION   U62622
VERSION     U62622.1  GI:1805253
KEYWORDS    .
SOURCE      cucumber.
  ORGANISM  Cucumis sativus
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Cucurbitales; Cucurbitaceae; Cucumis.
REFERENCE   1  (bases 1 to 2142)
  AUTHORS   Shimojima,M., Ohta,H., Iwamatsu,A., Masuda,T., Shioi,Y. and
            Takamiya,K.
  TITLE     Cloning of the gene for monogalactosyldiacylglycerol synthase and
            its evolutionary origin
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 94 (1), 333-337 (1997)
  MEDLINE   97144442
REFERENCE   2  (bases 1 to 2142)
  AUTHORS   Ohta,H., Shimojima,M., Iwamatsu,A., Masuda,T., Shioi,Y. and
            Takamiya,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1996) Bioscience & Biotechnology, Tokyo Institute
            of Technology, Nagatsuta 4259, Midori-ku, Yokohama 226, Japan
FEATURES             Location/Qualifiers
     source          1. .2142
                     /organism="Cucumis sativus"
                     /strain="Aonagajibai"
                     /db_xref="taxon:3659"
                     /dev_stage="5-day-old seedlings"
                     /note="illuminated for 6 h"
     CDS             317. .1894
                     /EC_number="2.4.1.46"
                     /function="glycosyltransferase;
                     UDP-galactose:1,2-diacylglycerol
                     3-beta-galactosyltransferase"
                     /codon_start=1
                     /product="monogalactosyldiacylglycerol synthase"
                     /protein_id="AAC49624.1"
                     /db_xref="GI:1805254"
                     /translation="MRNPSTVVQENGSVSDFISQLGYFAFSSRFLNLNSEGCSGSSSH
                     SLYLNGFENYRCVKRPPRSGASLSLSSSRGSSSLRRFVNEFNNVIKFHCHKPPLGFASL
                     GGVSDETNGIRDDGFGVSQDGALPLNKIEAENPKRVLILMSDTGGGHRASAEAIKAAF
                     NEEFGNNYQVFITDLWTDHTPWPFNQLPRSYNFLVKHGTLWKMTYYVTAPKVIHQSNF
                     AATSTFIAREVAKGLMKYRPDIIISVHPLMQHVPIRILRSKGLLNKIVFTTVVTDLST
                     CHPTWFHKLVTRCYCPSTEVAKRALTAGLQPSKLKVFGLPVRPSFVKPIRPKIELRKE
                     LGMDENLPAVLLMGGGEGMGPIEATAKALSKALYDENHGEPIGQVLVICGHNKKLAGR
                     LRSIDWKVPVQVKGFVTKMEECMGACDCIITKAGPGTIAEAMIRGLPIILNDYIAGQE
                     AGNVPYVVENGCGKFSKSPKEIANIVAKWFGPKADELLIMSQNALRLARPDAVFKIVH
                     DLHELVKQRSFVPQYSG"
BASE COUNT      597 a    406 c    489 g    650 t
ORIGIN
```

```
  Query Match             4.2%;  Score 48;  DB 14;  Length 2142;
  Best Local Similarity   51.9%;  Pred. No. 0.37;
  Matches  108;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy     776 aagttctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgatta 835
           | ||   ||| | |||       || || ||     |  |       |||||||| || |
Db    1524 AGGTGAAGGGGTTTGTCACAAAAATGGAGGAATGCATGGGAGCTTGTGATTGCATTATAA 1583

Qy     836 ccaagcccggcggcattactttgacagaagccacagccattggagtgcctgtcattctgt 895
           | ||| | |||      ||  |  | |||||||     |  ||  | ||| | ||||||
Db    1584 CAAAGGCGGGCCCTGGAACGATTGCTGAAGCCATGATAAGAGGTCTTCCTATAATTCTGA 1643

Qy     896 acaaacccgtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcggagctg 955
           |  |   | |  |||| ||||||   | |||||     |   | |  ||   ||||  ||
Db    1644 ATGACTACATTGCTGGACAGGAAGCTGGAAATGTGCCATATGTCGTCGAAAACGGATGTG 1703

Qy     956 ccatcgttgtgaaccgtcatgaagagat 983
           |   ||   ||   || | |||||||
Db    1704 GGAAGTTTTCAAAATCTCCTAAAGAGAT 1731
```

```
RESULT  13
AC026884/c
LOCUS       AC026884   150583 bp    DNA             PRI       10-NOV-2000
DEFINITION  Homo sapiens chromosome 10 clone RP11-123G9, complete sequence.
ACCESSION   AC026884
VERSION     AC026884.5  GI:11136696
KEYWORDS    HTG.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150583)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 150583)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   3  (bases 1 to 150583)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Nov 10, 2000 this sequence version replaced gi:9887582.
FEATURES             Location/Qualifiers
     source          1. .150583
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-123G9"
                     /clone_lib="RPCI-11"
BASE COUNT    48699 a  29525 c  28484 g  43875 t
ORIGIN


  Query Match              3.9%;  Score 44.4;  DB 88;  Length 150583;
  Best Local Similarity   58.2%;  Pred. No. 4.2;
  Matches   78;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy  -  154 gaggtaactcaatacctttatttaaaaagcttctcaatcgggaaacagttttatcgtttg 213
          ||||||  | | || |  || | |||||  | |   ||  ||||  |   |||| |
Db   64008 GAGGTATTTTACTAGCCATACTAAAAAACATACAATATTTGGAATTATAATTATTGAGAA 63949

Qy     214 ttttattacggagttgacaaaatctataataaacgtaaattcaatatttacttttaaaatg 273
           | |   |   |  |||||||| |||||| |    | |||  ||| ||||  ||||| |
Db   63948 ATATGGGATCTACCTGACAAAAACTATAAAATCTTTTAATAAAATTTTTATGTTAAATTA 63889

Qy     274 ggtaataaaagatt 287
             | |||||  |||
Db   63888 TTAACTAAAAAATT 63875


RESULT  14
AL354684/c
LOCUS       AL354684   170934 bp    DNA             HTG       23-JAN-2001
DEFINITION  Homo sapiens chromosome 13 clone RP11-191O21, *** SEQUENCING IN
            PROGRESS ***, 17 unordered pieces.
ACCESSION   AL354684
VERSION     AL354684.4  GI:9863681
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170934)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 21, 2000 this sequence version replaced gi:9213264.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA191O21
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 160941 bases at least Q40
            Consensus quality: 164931 bases at least Q30
            Consensus quality: 167496 bases at least Q20
            Insert size: 169334; sum-of-contigs
            Insert size: 164439; agarose-fp
            Quality coverage: 3.29x in Q20 bases; sum-of-contigs Quality
            coverage: 3.70x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    11172: contig of 11172 bp in length
            *    11173 11272: gap of      100 bp
            *    11273    14966: contig of 3694 bp in length
            *    14967 15066: gap of      100 bp
            *    15067    19307: contig of 4241 bp in length
            *    19308 19407: gap of      100 bp
            *    19408    63180: contig of 43773 bp in length
            *    63181 63280: gap of      100 bp
            *    63281    68271: contig of 4991 bp in length
            *    68272 68371: gap of      100 bp
            *    68372    70384: contig of 2013 bp in length
            *    70385 70484: gap of      100 bp
            *    70485    73907: contig of 3423 bp in length
            *    73908 74007: gap of      100 bp
            *    74008    81123: contig of 7116 bp in length
            *    81124 81223: gap of      100 bp
            *    81224    98475: contig of 17252 bp in length
            *    98476 98575: gap of      100 bp
            *    98576   101002: contig of 2427 bp in length
            *   101003 101102: gap of      100 bp
            *   101103   147781: contig of 46679 bp in length
            *   147782 147881: gap of      100 bp
            *   147882   150311: contig of 2430 bp in length
            *   150312 150411: gap of      100 bp
            *   150412   153588: contig of 3177 bp in length
            *   153589 153688: gap of      100 bp
            *   153689   157277: contig of 3589 bp in length
            *   157278 157377: gap of      100 bp
            *   157378   166456: contig of 9079 bp in length
            *   166457 166556: gap of      100 bp
            *   166557   168668: contig of 2112 bp in length
            *   168669 168768: gap of      100 bp
            *   168769   170934: contig of 2166 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170934
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-191O21"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .11172
                     /note="assembly_fragment:00531
                     fragment_chain:1"
     misc_feature    11273. .14966
                     /note="assembly_fragment:01529
                     fragment_chain:1"
     misc_feature    15067. .19307
                     /note="assembly_fragment:00958
                     fragment_chain:1"
     misc_feature    19408. .63180
                     /note="assembly_fragment:00538
                     fragment_chain:2"
```

```
     misc_feature    63281. .68271
                     /note="assembly_fragment:00676
                     fragment_chain:2"
     misc_feature    68372. .70384
                     /note="assembly_fragment:00628
                     fragment_chain:3"
     misc_feature    70485. .73907
                     /note="assembly_fragment:00839
                     fragment_chain:3"
     misc_feature    74008. .81123
                     /note="assembly_fragment:01002
                     fragment_chain:4"
     misc_feature    81224. .98475
                     /note="assembly_fragment:01050
                     fragment_chain:4"
     misc_feature    98576. .101002
                     /note="assembly_fragment:01493
                     fragment_chain:5"
     misc_feature    101103. .147781
                     /note="assembly_fragment:00305
                     fragment_chain:5"
     misc_feature    147882. .150311
                     /note="assembly_fragment:00195"
     misc_feature    150412. .153588
                     /note="assembly_fragment:00504"
     misc_feature    153689. .157277
                     /note="assembly_fragment:00743"
     misc_feature    157378. .166456
                     /note="assembly_fragment:00830"
     misc_feature    166557. .168668
                     /note="assembly_fragment:01170"
     misc_feature    168769. .170934
                     /note="assembly_fragment:01474"
BASE-COUNT    53858 a  33064 c  31911 g  50497 t   1604 others
ORIGIN


  Query Match             3.9%;  Score 44.4;  DB 79;  Length 170934;
  Best Local Similarity   58.2%;  Pred. No. 4.3;
  Matches   78;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy     154 gaggtaactcaatacctttatttaaaaagcttctcaatcgggaaacagttttatcgtttg 213
           ||||||  | | || |  || | |||||  | |   ||  ||||  |   |||| |
Db   65107 GAGGTATTTTACTAGCCATACTAAAAAACATACAATATTTGGAATTATAATTATTGAGAA 65048

Qy     214 ttttattacggagttgacaaaatctataataaacgtaaattcaatatttacttttaaaatg 273
           | |   |   |  ||||||||| |||||| |    | |||  ||| ||||  ||||| |
Db   65047 ATATGGGATCTACCTGACAAAAACTATAAAATCTTTTAATAAAATTTTTATGTTAAATTA 64988

Qy     274 ggtaataaaagatt 287
             | ||||| |||
Db   64987 TTAACTAAAAAATT 64974


RESULT 15
BSAJ2571/c
LOCUS       BSAJ2571    55593 bp    DNA             BCT       18-NOV-1997
DEFINITION  Bacillus subtilis 168 56 kb DNA fragment between xlyA and ykoR.
ACCESSION   AJ002571
VERSION     AJ002571.1  GI:2632001
KEYWORDS    dppA gene; dppB gene; dppC gene; dppD gene; dppE gene;
            gamma-glutamyl kinase; gamma-glutamyl phosphate reductase;
            intracellular serine protease; MetC; N-acetylmuramoyl-L-alanine
            amidase; tnrA gene; ykaA gene; ykaB gene; ykaC gene; ykaD gene;
            ykbA gene; ykcA gene; ykcB gene; ykcC gene; ykdA gene; ykeA gene;
            ykfA gene; ykfB gene; ykfC gene; ykfD gene; ykgA gene; ykgB gene;
            ykhA gene; ykiA gene; ykjA gene; ykkA gene; ykkB gene; ykkC gene;
            ykkD gene; ykkE gene; yklA gene; ykmA gene; yknA gene; ykoA gene;
            ykoB gene; ykoC gene; ykoD gene; ykoE gene; ykoF gene; ykoG gene;
            ykoH gene; ykoI gene; ykoJ gene; ykoK gene; ykoL gene; ykoM gene;
            ykoN gene; ykoP gene; ykoQ gene; ykoR gene.
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 55593)
  AUTHORS   Devine,K.M.
  TITLE     Sequence of the Bacillus subtilis genome between xlyA and ykoR
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 55593)
  AUTHORS   Devine,K.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1997) Kevin M. Devine, Genetics, Trinity College
            Dublin, Lincoln Place Gate, Dublin, 2, Ireland
FEATURES             Location/Qualifiers
     source          1. .55593
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            1. .314
                     /gene="xlyA"
     CDS             <1. .314
                     /gene="xlyA"
                     /function="PBSX autolysin"
                     /codon_start=3
                     /transl_table=11
                     /product="N-Acetylmuramoyl-L-alanine amidase"
                     /protein_id="CAA05561.1"
                     /db_xref="GI:2632002"
                     /db_xref="SWISS-PROT:P39800"
                     /translation="IRVGQVLIVSAPSAAEKPELYPLPDGIIQLTTPYTSGEHVFQVQ
                     RALAALYFYPDKGAVNNGIDGVYGPKTADAVARFQSVNGLTADGIYGPATKEKIAAQL
                     S"
     gene            complement(574. .744)
                     /gene="ykaD"
     CDS             complement(574. .744)
                     /gene="ykaD"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="YkaD"
                     /protein_id="CAA05562.1"
                     /db_xref="GI:2632003"
                     /db_xref="SPTREMBL:O34800"
                     /translation="MERAFQNRCEPRAAKPFKILKKRSTTSVASYQVSPHTARIFKEN
                     ERLIDEYKRKKA"
     gene            complement(744. .1490)
                     /gene="ykaC"
     CDS             complement(744. .1490)
                     /gene="ykaC"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="YkaC"
                     /protein_id="CAA05563.1"
                     /db_xref="GI:2632004"
                     /db_xref="SPTREMBL:O34853"
                     /translation="MVLFFQIMVWCIVAGLGLYVYATWRFEAKVKEKMSAIRKTWYLL
                     FVLGAMVYWTYEPTSLFTHWERYLIVAVSFALIDAFIFLSAYVKKLAGSELETDTREI
                     LEENNEMLHMYLNRLKTYQYLLKNEPIHVYYGSIDAYAEGIDKLLKTYADKMNLTASL
                     CHYSTQADKDRLTEHMDDPADVQTRLDRKDVYYDQYGKVVLIPFTIETQNYVIKLTSD
                     SIVTEFDYLLFTSLTSIYDLVLPIEEEGEG"
     gene            complement(1614. .2600)
                     /gene="ykaB"
     CDS             complement(1614. .2600)
                     /gene="ykaB"
                     /note="homologous to low affinity phosphate transporter"
                     /codon_start=1
                     /transl_table=11
                     /product="YkaB"
                     /protein_id="CAA05564.1"
                     /db_xref="GI:2632005"
                     /db_xref="SPTREMBL:O34436"
                     /translation="MDILLILTILIVICALAFDFINGFHDTANAIATSVSTKALKPRH
                     AIILAAVMNFVGAMTFTGVAKTITKDIVDPYTLENGSVVILAALLAAIAWNLITWYYG
                     IPSSSSHAIIGAIAGAAIAAAGFAALNYKGFIKIIEALILSPIIAFVLGFILYSIVKL
```

IFKDSNLAKTNKQFRRVQIVTAALQSYTHGTNDAQKAMGIITMALITANLHTSANDIP
TWVQFACATAMGLGTSIGGWKIIKTVGGKIMKIRPVNGVSADLTGAAIIFGATFIHLP
VSTTHVISSSILGVGASHRVKGVNWGTAKRISLHGSSRFRFQRHLVPSPTLF"

gene complement(2613. .3230)
/gene="ykaA"

CDS complement(2613. .3230)
/gene="ykaA"
/function="unknown"
/codon_start=1
/transl_table=11
/product="YkaA"
/protein_id="CAA05565.1"
/db_xref="GI:2632006"
/db_xref="SPTREMBL:O34454"
/translation="MIRRKKDKFSLLLTEIAKNIDETAEYFVNFKVTNQTTLKEFADT
LKEYETKGDNHVHVMIKELNKAFITPIEREDILQLTNSLDDVLDGIEHFSAMMEIFSI
TSSDEHIDKFSGYIRECAKEILITIELLAENRLKDIQPHAIKIKEYEHSCDNLHRKSL
KNLFGNETDPIKVIQYREIYETLEEIADSCQSVANNLETIIMKNA"

gene complement(3506. .4822)
/gene="ykbA"

CDS complement(3506. .4822)
/gene="ykbA"
/function="unknown"
/codon_start=1
/transl_table=11
/product="YkbA"
/protein_id="CAA05566.1"
/db_xref="GI:2632007"
/db_xref="SPTREMBL:O34739"
/translation="MHTEDNGLKKEIGLLFALTLVIGTIIGSGVFMKPGAVLAYSGDS
KMALFAWLLGGILTLAGGLTIAEIGTQIPKTGGLYTYLEEVYGEFWGFLCGWVQIIIY
GPAIIGALGLYFGSLMANLFGWGSGLSKVIGIIAVLFLCVINIIGTKYGGFVQTLTTI
GKLIPIACIIVFGLWKGDQHIFTAVNESISDMNFGAAILATLFAYDGWILLAALGGEM
KNPEKLLPRAMTGGLLIVTAIYIFINFALLHILSANEIVTLGENATSTAATMLFGSIG
GKLISVGIIVSIFGCLNGKVLSFPRVSFAMAERKQLPFAEKLSHVHPSFRTPWIAISF
QIALALIMMLISNPDKLSEISIFMIYIFYVMAFFAVFILRKRAKGEKRAYSVPLYPFM
PILAIAGSFFVLGSTLITDTMSCGLSILIGLAGLPVYYGMKKRKAS"

gene 5211. .6161
/gene="ykcA"

CDS 5211. .6161
/gene="ykcA"
/note="putative dioxygenase"
/codon_start=1
/transl_table=11
/product="YkcA"
/protein_id="CAA05567.1"
/db_xref="GI:2632008"
/db_xref="SPTREMBL:O34689"
/translation="MKVNGIHHVSALTADAQKNLDFYKKVLGLKLVKKSVNQDEPTMY
HLFYGDEVANPGTELTFFEIPRIAPFHAGTNSISSIGLRVPGTEALHYWKERFEEQQV
THSGISKRAGRDILAFQDHEGQRLVLTADEEGKGYGLPVKQSGIPEEFSFRGLGPVEL
TVPYAEPTLHVLTNILGFTEISREPVEGQGTAVILESGEGGAATEIHLIERNDLPRER
QGKGSVHHVAFRVRDEEELAGWHRIISREGFSNSGIVERYYFKALYFREPNGILFELS
TDGPGFMVDENLDELGQTIALPPYLEHRRAEIEAKLKPIQ"

gene 6412. .8562
/gene="ykcB"

CDS 6412. .8562
/gene="ykcB"
/function="unknown"
/codon_start=1
/transl_table=11
/product="YkcB"
/protein_id="CAA05568.1"
/db_xref="GI:2632009"
/db_xref="SPTREMBL:O34575"
/translation="MEKKKRELDIVLLLILLASAFLNIYNIWKDDTVNPYYTAAVTSM
MQSFHNFFYASFDAAGFITVDKPPVTYQIQTISALIFGMHGWSVILPQALAGVGSVLL
MYLLIKPTFGKTAARIASFVMACTPIAVAVARTNNVDALLVFFLLLATWLLFKAIRKG
KLIWLLAAFFVVGVGFNTKMLQAYMILPAFLLFYLIAANATIKKKIVSLVSALAVLAA
VSLSWPLIVDNIPASKRPYIGSSQTNSVLELAFGYNGIQRLTGQNSGGGQGAPNKDAS
KEMSSSDNTQAPPNQSSSNSSSSDGKSSNGNMAAPPSNGQMPSGGQGGPPSGGDGGQG
GPGGDGGKGGTGTGSKMQSGSGMFGTGTPGPLRLFQQELSDQISWLLPFAIFGIAGLL
IAGARERRRLSVEQKETVFWVAWLVPIAGFFSVAEFFHHYYLIMLAPRFAPLVGAGWV
ALVHLYRNQTGWKAWLLPGAIIATTGFELFILRNYNDQIGVGWSIGVGVIGVLSAIAL
LLFKQRQKPFSYYVSLAALLALLVMPMYWASTPLLYGGNSSLPETGPQLASMSGKGMG
MSDATVNEKLIKYLEENNSGAEYLFATTDSNTAAPYIIKTKKAVMAIGGYSGSDPAIT
LTQFKKLVKEGKVKYFLASGMGRGGNNDIVEWVEKNGKEVASEKWQSSSDQKTENTDS
ADTSSSKASGENGKMGGPGGMNQSATLYELHADE"

gene 8574. .9545
/gene="ykcC"

CDS 8574. .9545
/gene="ykcC"
/note="putative dolichol phosphate mannose synthetase"
/codon_start=1
/transl_table=11
/product="YkcC"
/protein_id="CAA05569.1"
/db_xref="GI:2632010"
/db_xref="SWISS-PROT:O34319"
/translation="MSRHIQYSIVVPVYNEELVIHETYQRLKEVMDQTKENYELLFVN
DGSKDRSIEILREHSLIDPRVKIIDFSRNFGHQIAITAGMDYAQGNAIVVIDADLQDP
PELILEMIEKWKEGYEVVYAVRTKRKGETFFKKQTAAMFYRLLSGMTDIDIPIDTGDF
RLMDRKVCDEMKRLKEKNPFVRGLVSWVGSKQTAVEYVRDERLAGETKYPLKKMLKLS
MDGITTFSHKPLKLASYAGILMSGTGFLYMFIVLYLKLFTDSTITGWSSLIVIQLLFS
GIVLLILGVIGEYIGRIYDEAKDRPLYIVQKSYGIENKRLYRDQHMS"

gene complement(10063. .11412)
/gene="ykdA"

CDS complement(10063. .11412)
/gene="ykdA"
/note="putative serine protease, heat-shock inducible;
homologous to HtrA from E. coli"
/codon_start=1
/transl_table=11
/product="YkdA"
/protein_id="CAA05570.1"
/db_xref="GI:2632011"
/db_xref="SPTREMBL:O34358"
/translation="MDNYRDENRTKGNENEVFLTKENDQSASYSARNVIHDQEKKKRG
FGWFRPLLGGVIGGSLALGIYTFTPLGNHDSQDTAKQSSSQQQTQSVTATSTSSESKK
SSSSSSAFKSEDSSKISDMVEDLSPAIVGITNLQAQSNSSLFGSSSSDSSEDTESGSG

Query Match 3.8%; Score 44.2; DB 2; Length 55593;
Best Local Similarity 54.7%; Pred. No. 4.3;
Matches 88; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```
Qy     392 tcaacgttatgactgatttttgtcttcataaaatttgggttcacgaaaacgtggataaaat 451
           | || ||    ||||||||||   |  || || | ||||  |   | ||| | ||| |
Db   51832 TGAATGTATACACTGATTTTTTCGTGAATCAATTATGGGGACGGAAGAACATTGATTACC 51773

Qy     452 attatgtggcgacagattacgtgaaggaaaaactgctggagatcggcactcatccaagca 511
           ||| |||  |||      | || ||| |  | ||  |       || |   |||  | ||
Db   51772 ATTTTGTTCCGAGCACGGAAGTAAAGAAGCAGCTAATATCAGAGGGTATCGATCAGAACA 51713

Qy     512 atgtaaaaatcacaggaattccaatcaggccgcaatttgaa 552
           |  || |  | || || |||||| |    |   | ||||||
Db   51712 ACATATACTTGACCGGCATTCCAGTTCACCAAAATTTTGAA 51672
```

Search completed: June 29, 2001, 11:53:36
Job time: 8429 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 29, 2001, 08:57:18 ; Search time 28.15 Seconds
(without alignments)
1033.701 Million cell updates/sec

Title: US-09-668-788-2
Perfect score: 1970
Sequence: 1 MNTNKRVLILTANYGNGHVQ.....EDILKESEMMTAKQKAKVLS 382

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1970	100.0	382	2	C69935	cell wall synthesi
2	368	18.7	374	2	E83894	hypothetical prote
3	353	17.9	411	2	F75439	probable cell wall
4	317	16.1	464	2	C84499	probable monogalac
5	309	15.7	373	2	B69860	conserved hypothet
6	303.5	15.4	525	2	T10478	probable 1,2-diacy
7	301.5	15.3	533	2	T05092	probable 1,2-diacy
8	293	14.9	468	2	T52269	1,2-diacylglycerol
9	196	9.9	363	2	JC1275	phospho-N-acetylmu
10	195	9.9	344	2	C70401	phospho-N-acetylmu
11	178	9.0	363	2	E83970	UDP-N-acetylglucos
12	171	8.7	351	2	D64185	UDP-N-acetylglucos
13	170.5	8.7	363	2	F70195	UDP-N-acetylglucos
14	163	8.3	383	2	E70156	lipopolysaccharide
15	159	8.1	357	2	E86823	peptidoglycan synt
16	156.5	7.9	398	2	F64456	hypothetical prote
17	138	7.0	378	2	A72304	UDP-N-acetylglucos
18	137.5	7.0	364	2	T34954	probable UDP-N-ace
19	136.5	6.9	458	2	C71420	hypothetical prote
20	136	6.9	353	2	C64664	transferase, pepti
21	135	6.9	484	2	E64432	spore coat polysac
22	134	6.8	353	2	G71852	udp-n-acetylglucos
23	130.5	6.6	392	1	C69851	macrolide glycosyl
24	128.5	6.5	410	2	E70579	probable murG prot
25	128	6.5	406	2	C72340	probable hexosyltr
26	123.5	6.3	367	2	D82763	UDP-N-acetylglucos
27	123	6.2	355	1	BVECMG	UDP-N-acetylglucos
28	123	6.2	355	2	F85491	hypothetical prote
29	122.5	6.2	339	2	B72402	UDP-N-acetylglucos
30	122.5	6.2	342	2	D81306	probable UDP-N-ace
31	121.5	6.2	385	2	C71699	murg protein (murG
32	119	6.0	418	2	G75496	UDP-N-acetylglucos
33	119	6.0	478	2	T08395	UTP-glucose glucos
34	117.5	6.0	380	2	D70049	UDP-N-acetylglucos
35	117	5.9	400	1	F69142	probable hexosyltr
36	116.5	5.9	338	2	E71699	capm protein (capM
37	116.5	5.9	366	1	D70351	probable hexosyltr
38	116	5.9	363	1	C72590	probable hexosyltr
39	116	5.9	379	2	E86651	LPS biosynthesis p
40	115.5	5.9	571	2	B64469	hypothetical prote
41	114.5	5.8	362	2	B86783	UDP-N-acetylglucos
42	114.5	5.8	404	1	S77553	probable hexosyltr
43	113.5	5.8	354	2	D84955	hypothetical prote
44	113	5.7	371	2	S76863	hypothetical prote
45	112.5	5.7	1068	1	JQ1329	sucrose-phosphate

ALIGNMENTS

RESULT 1
C69935
cell wall synthesis homolog ypfP - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: C69935
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033
A;Accession: C69935
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-382 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14110.1; PID:e11836
A;Experimental source: strain 168
C;Genetics:
A;Gene: ypfP

Query Match 100.0%; Score 1970; DB 2; Length 382;
Best Local Similarity 100.0%; Pred. No. 2.4e-127;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MNTNKRVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MNTNKRVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLK 60

Qy      61 SFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 SFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEY 120

Qy     121 RRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKITGIPIR 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 RRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKITGIPIR 180

Qy     181 PQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVVVCGKN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 PQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVVVCGKN 240
```

```
Qy     241 TALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGITLTEATAIGVPVILYK 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGITLTEATAIGVPVILYK 300

Qy     301 PVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDTLHRMKKNIKDLHLANSSEV 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 PVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDTLHRMKKNIKDLHLANSSEV 360

Qy     361 ILEDILKESEMMTAKQKAKVLS 382
           ||||||||||||||||||||||
Db     361 ILEDILKESEMMTAKQKAKVLS 382
```

RESULT 2
E83894
hypothetical protein BH1957 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C;Accession: E83894
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20263314
A;Accession: E83894
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-374 <STO>
A;Cross-references: GB:AP001513; GB:BA000004; NID:g10174345; PIDN:BAB05676.1; GSPDB:GN00
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH1957

Query Match 18.7%; Score 368; DB 2; Length 374;
Best Local Similarity 27.0%; Pred. No. 7.2e-18;
Matches 103; Conservative 84; Mismatches 159; Indels 36; Gaps 13;

```
Qy       8 LILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLKSFSIGKQ 67
           || :|: |||| | || |  :    |:|   :   |  | | : :     |:    :| :
Db       6 LIFSASIGNGHNQAAKALQVEFQNKGYQPEIIDTFYSLS-PALHKFMLTSYVNLLKVGPR 64

Qy      68 FYRLFYYGVDK---IYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEYRRRT 124
            ::  |:. :|        :|  :|    : |   |  ::   :::| | :     | ::
Db      65 IWQKIYFQAEKYPLFLFLDQFATFFV---ESLHATVKSNRCSFLVSTHPFVTAFLVRLKS 121

Qy     125 GRV--IPTFNVMTDFCLHKIWVHENVDKYYVA----TDYVKEKLLEIGTHPSNVK----- 173
            :   :| : |:||| ||  ::   :| |: :    ||: |          :||
Db     122 KKQLNLPLYTVITDFVLHPAYLRPEIDGYFTSDPNFTDFAK---------LNNVSDDRFF 172

Qy     174 ITGIPIRPQFEE-SMPVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENL-VKDDQV 231
           |||||  |   |    |   :     | | :||||:|  |  | | |  ::   |    : :
Db     173 PTGIPI-PNLESIDQPKWKVRNDLGLDPQRKVLIIAGGGIG-LTNYAQVIRALECLPEPI 230

Qy     232 QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGITLTEATA 291
           |:: : | |  :|| :| ::::  :|||: : ::     : :| :::| ||:|: |:
Db     231 QLLCMIGHNYQVKEKISRIKSKH--ELKVIEFTDKFLLYLKASDAILSKAGGLTMAESLV 288

Qy     292 IGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDTLHRMKKNIKD 351
           |:|:::|||| |: || |  | |||: |   :|| ::  :| :|     | :| :
Db     289 CETPIIIHQPVPGHEEHNAKFLIDAGAALRVKGSKEIPTTIKRVLYEEACFGPMIENARK 348

Qy     352 LHLANSSEVILEDIL---KESE 370
           |   |::  |:| :|   || :
Db     349 LKKPNAANEIVEQMLLLVKEQQ 370
```

RESULT 3
F75439
probable cell wall synthesis protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: F75439
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: F75439
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-411 <WHI>
A;Cross-references: GB:AE001958; GB:AE000513; NID:g6458805; PIDN:AAF10649.1; PID:g645
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR1076
A;Map position: 1

Query Match 17.9%; Score 353; DB 2; Length 411;
Best Local Similarity 26.2%; Pred. No. 8.7e-17;
Matches 101; Conservative 77; Mismatches 164; Indels 44; Gaps 11;

```
Qy       6 RVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYL--YLKSFS 63
           | | :: : | || |       :| |:  |    |  |  | : :     :||  : :||:
Db      29 RALFMSVSLGAGHDQA-----QQAVKQAFAERGVELLGAEHDSV-----EYLSTFERSFT 78

Qy      64 I---------GKQFYRLFYYGVDKIYNKRKFNIYFKM----GNKRLGELVDEHQPDIIIN 110
           :              || ||:  |:    : :||  :|    |     : : | :|:::||
Db      79 VDLYEFELRYAPWLYRGFYWLTDQ---DQPWNIISRMFTWLGMGAFKDELRELRPEVVIN 135

Qy     111 TF--PMIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIGTH 168
           :|  |  |    | :||:      ::||:  |  |     |   ||::  : ::|| |
Db     136 SFWAPAAVCDTLRAQTGQRFLNCLIVTDYRAHLHWARRETDLLMVASEETRRQMLERGVR 195

Qy     169 PSNVKITGIPIRPQFEESMPV------GPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELC 222
           |  |::||||| | | | :          :: : :| |   :||:  |  |      ::
Db     196 PEQVEVTGIPISPAFREVLAADRWALRAELFSEMSLRPGVPLLLLSGGGRGHYAAAADVL 255

Qy     223 ENLVK-DDQVQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKP 281
           |      |||:|   :     |:        |  :  ||:   :  |   :| :: |
Db     256 TELGNLGRAVQVLVPASRQGEGTETI------GGATVHHLGFRRDLPRLLAASDLVVGKA 309

Qy     282 GGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDT 341
           ||:|: ||||:|||:::| |:|||||: ||:| |  || :      ::   |   | |
Db     310 GGLTVAEATALGVPLVIYAPIPGQEEHNADFLERHGAGLWARARHDVRPLVLRAL-DPAE 368

Qy     342 LHRMKKNIKDLHLANSSEVILEDILK 367
           |:    : : : :::: :   ||:
Db     369 HARLSAGARAVGIPDAADRVAGAILR 394
```

RESULT 4
C84499
probable monogalactosyldiacylglycerol synthase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84499
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: C84499
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-464 <STO>
A;Cross-references: GB:AE002093; NID:g4734000; PIDN:AAD28678.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g11810
A;Map position: 2

Query Match 16.1%; Score 317; DB 2; Length 464;
Best Local Similarity 25.6%; Pred. No. 3e-14;

```
  Matches  101;  Conservative   90;  Mismatches  171;  Indels   32;  Gaps   16;

Qy        5 KRVLILTANYGNGHVQVAKTLYEQCVRLGF---QHVTVSNLYQESN--PIVSEVTQYLYL 59
              | ||||  :: |  ||    |: :       :: |     : : ::::|      |:        || ::
Db       72 KTVLILMSDTGGGHRASAEAI-RDAFKIEFGDDYRIIIKDVWKEYTGWPLNDMERQYKFM 130

Qy       60 KSFSIGKQFYRLFYYGVD-KIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM---- 114
                  :|    : : ::|    |   :|    :            |  :    : |::|||||:  |:
Db      131 VK-HVG--LWSVAFHGTSPKWIHKSYLSALAAYYAKEIEAGLMEYKPDIIISVHPLMQHI 187

Qy      115 -IVVPEYRRRTGRVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNV 172
              :  |  :::       :||      |:||     |:  |   |    |   :    |   |  ::  |      |  :
Db      188 PLWVMKWQGLHKKVI-FVTVITDLNTCHRTWFHHGVSRCYCPSKEVAKRALVDGLDDSQI 246

Qy      173 KITGIPIRPQFEES-MPVGPIYKKYNLSPNKKVLLIMAGAHG---VLKNVKELCENLVKD 228
              :: |:|:||  |   :  :        : |:   :    |      :|:|  |    |       | |       | ::|
Db      247 RVFGLPVRPSFPRTILNKNELRKELEIDLNLPAVLLMGGGEGMGPVQKTALALGDSLYNS 306

Qy      229 DQV----QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGI 284
               :          |::|:||:|    |    :|::  |  :        :||  |:   ::::         ||:||| |
Db      307 KESNPIGQLIVICGRNKVLASTLASHEWK--IPVKVRGFETQMEKWMGACDCIITKAGPG 364

Qy      285 TLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLA-DEDTLH 343
              |: ||     |:|:||      :|||||  |     :    | ||  :           :|    :  |           :  :::  |
Db      365 TIAEALICGLPIILNDYIPGQEKGNVPYVVDNGAGVFTRSPKETAKIVADWFSNNKEELK 424

Qy      344 RMKKNIKDLHLANSSEVILEDILKESEMMTAKQK 377
              :|  :|    |  |:      |    ||:|:      :: :|:
Db      425 KMSENA--LKLSQPEAVF--DIVKDIHHLSQQQQ 454
```

RESULT 5
B69860
conserved hypothetical protein ykoN - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: B69860
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauee
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033
A;Accession: B69860
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-373 <KUN>
A;Cross-references: GB:Z99110; GB:Z99111; GB:AL009126; NID:g2633699; PIDN:CAB13208.1; PI
A;Experimental source: strain 168
C;Genetics:
A;Gene: ykoN

```
  Query Match             15.7%;  Score 309;  DB 2;  Length 373;
  Best Local Similarity   24.6%;  Pred. No. 7.7e-14;
  Matches   99;  Conservative   89;  Mismatches  158;  Indels   56;  Gaps   16;

Qy        5 KRVLILT-ANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLKSFS 63
              | :||    :    ||  ||   |   :     ||        : :::   |     : :::     |||
Db        2 KNILIFPFLSISTGHHHVADALQAELESQGLAAEKI-DIFSHSYRRLEKLSSVAYLKWI- 59

Qy       64 IGKQFYRLFYYGVDKI-------YNKRKFNIYFKMGNKRLGELVDEHQPDIIINT--FPM 114
               |::    | |: ::          ::|| | :|   :   :::    :: |  ||||    |    |
```

```
Db       60 ---QYFPKTYSGIYRLLACGEFQHDKRYF-MYECVFTQQMRHILQEKQPDIAFCTHALPS 115

Qy      115 IVV----PEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIGTHPS 170
              ::       |||     |             || ||| ::::|  :|:| ::| :  ||::|:   |    :
Db      116 YLLNRLKPEYPNLT-----VVNVYTDFFVNQLWGRKNIDYHFVPSTEVKKQLISEGIDQN 170

Qy      171 NVKITGIPIRPQFE----ESMPVGPIYKKYNLSPNKKVLLIMAGAHGV---LKNVKELCE 223
              |: :||||:   ||      :::    | |                  ::|   |: ||     || |:||
Db      171 NIYLTGIPVHQNFEMESADTLQHHPPY----------TIIITGGSMGVGGILKWVQELSP 220

Qy      224 NLVKDDQVQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGG 283
                        ::    ::||:|    |        :  :|          :  :    |          :: |:          ::|||||
Db      221 G----GKILYKILCGRNEKLYSYVKSLHHPLIEAIPYLHSKAEMNRLYEQATGIMTKPGG 276

Qy      284 ITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRH----EEILESVTSLLADE 339
              |:|::|         :||  :|   :||||: | |.  :|      :|  :            ::   | :   :
Db      277 VTISECLQKRLPVFIYHALPGQEEMNLNLLHER--KLVTDMRNWDMQKAEEYIAAFFQSN 334

Qy      340 DTLHRMKKNIKDLHLANSSEVILEDILKESEMMTAKQKAKVL 381
              :  :     ||::    :|     |:   ::|:||      :    |||    :|
Db      335 EQMKEYKKHVNG-YLGEMSDRKIKDVLKR---IIWKQKNTLL 372
```

RESULT 6
T10478
probable 1,2-diacylglycerol 3-beta-galactosyltransferase (EC 2.4.1.46) precursor, chl
N;Alternate names: monogalactosyldiacylglycerol synthase
C;Species: Cucumis sativus (cucumber)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T10478
R;Shimojima, M.; Ohta, H.; Iwamatsu, A.; Masuda, T.; Shioi, Y.; Takamiya, K.
Proc. Natl. Acad. Sci. U.S.A. 94, 333-337, 1997
A;Title: Cloning of the gene for monogalactosyldiacylglycerol synthase and its evolut
A;Reference number: Z17042; MUID:97144442
A;Accession: T10478
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-525 <SHI>
A;Cross-references: EMBL:U62622; NID:g1805253; PIDN:AAC49624.1; PID:g1805254
A;Experimental source: cv. Aonagajibai; 5 day old seedlings
C;Genetics:
A;Genome: nuclear
C;Function:
A;Description: catalyzes the formation of monogalactosyldiacylglycerol, a major struc
A;Pathway: membrane lipid biosynthesis
C;Keywords: chloroplast; glycosyltransferase; hexosyltransferase
F;1-103/Domain: transit peptide (chloroplast) #status predicted <TNP>
F;104-525/Product: 1,2-diacylglycerol 3-beta-galactosyltransferase #status predicted

```
  Query Match             15.4%;  Score 303.5;  DB 2;  Length 525;
  Best Local Similarity   25.7%;  Pred. No. 2.9e-13;
  Matches  100;  Conservative   77;  Mismatches  179;  Indels   33;  Gaps   13;

Qy        5 KRVLILTANYGNGH---VQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLKS 61
              |||||| :: | ||          :  |    :  :          :| |  :::|: :    |             |
Db      136 KRVLILMSDTGGGHRASAEAIKAAFNEEFGNNYQ-VFITDLWTDHTPWPFNQLPRSYNFL 194

Qy       62 FSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM-----IV 116
               |     :     :|      |: :: |                :  :  :  : :::|||||:   |:        |
Db      195 VKHGTLWKMTYYVTAPKVIHQSNFAATSTFIAREVAKGLMKYRPDIIISVHPLMQHVPIR 254

Qy      117 VPEYRRRTGRVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKIT 175
              :     :       ::: |   |:||     |   | |:  |  :  |    :   | ::  |    |    || :|:
Db      255 ILRSKGLLNKIVFT-TVVTDLSTCHPTWFHKLVTRCYCPSTEVAKRALTAGLQPSKLKVF 313

Qy      176 GIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGAHG---VLKNVKELCENLVKDD 229
              |:|:||  |  :     |:   |      :  |:     :      :|:|  |     |         :          |  |  :  |      ::
Db      314 GLPVRPSFVK--PIRPKIELRKELGMDENLPAVLLMGGGEGMGPIEATAKALSKALYDEN 371

Qy      230 Q----VQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGIT 285
                      |:|:|:||   |      |     | ::: :          ::|  |:|  :::|           ||:|||    |    |
Db      372 HGEPIGQVLVICGHNKKLAGRLRSIDWK--VPVQVKGFVTKMEECMGACDCIITKAGPGT 429
```

```
Qy    286 LTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADE-DTLHR 344
Db    430 IAEAMIRGLPIILNDYIAGQEAGNVPYVVVENGCGKFSKSPKEIANIVAKWFGPKADELLI 489

Qy    345 MKKNIKDLHLANSSEVI-----LEDILKE 368
Db    490 MSQNA--LRLARPDAVFKIVHDLHELVKQ 516
```

RESULT 7
T05092
probable 1,2-diacylglycerol 3-beta-galactosyltransferase (EC 2.4.1.46) - Arabidopsis tha
N;Alternate names: protein F28M20.30
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T05092
R;Bevan, M.; Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Hoheisel, J.; Mewes,
submitted to the Protein Sequence Database, November 1998
A;Reference number: Z15398
A;Accession: T05092
A;Molecule type: DNA
A;Residues: 1-533 <BEV>
A;Cross-references: EMBL:AL031004
A;Experimental source: cultivar Columbia; BAC clone F28M20
C;Genetics:
A;Map position: 4
A;Introns: 175/3; 233/2; 287/2; 334/3; 409/3; 430/3; 455/3
A;Note: F28M20.30
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 15.3%; Score 301.5; DB 2; Length 533;
Best Local Similarity 24.8%; Pred. No. 4.1e-13;
Matches 101; Conservative 86; Mismatches 172; Indels 49; Gaps 17;

```
Qy      5 KRVLILTANYGNGHVQVAKTLYEQCVRLGFQ-------HVTVSNLYQESN--PIVSEVTQ 55
Db    142 KKVLILMSDTGGGHRASA-----EAIRAAFNQEFGDEYQVFITDLWTDHTPWPFNQLPRS 196

Qy     56 YLYLKSFSIGKQFYRLFYYGVD-KIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM 114
Db    197 YNFLVKHG---TLWKMTYYGTSPRIVHQSNFAATSTFIAREIAQGLMKYQPDIIISVHPL 253

Qy    115 I--VVPEYRRRTG--RVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHP 169
Db    254 MQHVPLRVLRSKGLLKKIVFTTVITDLSTCHPTWFHKLVTRCYCPSTEVAKRAQKAGLET 313

Qy    170 SNVKITGIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGAHGV--LKNVKELCEN 224
Db    314 SQIKVYGLPVRPSFVK--PVRPKVELRRELGMDENLPAVLLMGGGEGMGPIEATARALAD 371

Qy    225 LVKDDQV-----QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMIT 279
Db    372 ALYDKNLGEAVGQVLIICGRNKKLQSKLSSLDWK--IPVQVKGFITKMEECMGACDCIIT 429

Qy    280 KPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLAD- 338
Db    430 KAGPGTIAEAMIRGLPIILNGYIAGQEAGNVPYVVENGC----GKFSKSPKEISKIVADW 485

Qy    339 ----EDTLHRMKKNIKDLHLANSSEVILEDILKESEMMTAKQKAKVLS 382
Db    486 FGPASKELEIMSQNA--LRLA-KPEAVFKIVHDMHELVRKKNSLPQLS 530
```

RESULT 8
T52269
1,2-diacylglycerol 3-beta-galactosyltransferase (EC 2.4.1.46) [imported] - Arabidopsis t
N;Alternate names: monogalactosyldiacylglycerol synthase
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C;Accession: T52269
R;Awai, K.; Shimojima, M.; Masuda, T.; Takamiya, K.I.; Ohta, H.
submitted to the EMBL Data Library, July 1998
A;Description: cDNA cloning of an Arabidopsis monogalactosyldiacylglycerol synthase r
A;Reference number: Z26011
A;Accession: T52269
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-468 <AWA>
A;Cross-references: EMBL:AJ000331; PIDN:CAA04005.1
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: mgd
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 14.9%; Score 293; DB 2; Length 468;
Best Local Similarity 25.1%; Pred. No. 1.3e-12;
Matches 98; Conservative 86; Mismatches 172; Indels 34; Gaps 14;

```
Qy      5 KRVLILTANYGNGHVQVAKTLYEQC-VRLGFQH-VTVSNLYQESN--PIVSEVTQYLYLK 60
Db     69 KNVLILMSDTGGGHRASAEAIRDAFKIEFGDKYRVIVKDVWKEYTGWPLNDMERSYKFMV 128

Qy     61 SFSIGKQFYRLFYYG-----VDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM- 114
Db    129 KH---VQLWKVAFHSTSPKWIHSCYLAAIAAYYAKEVEAGL----MEYKPEIIISVHPLM 181

Qy    115 ----IVVPEYRRRTGRVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHP 169
Db    182 QHIPLWVLKWQELQKRVL-FVTVITDLNTCHPTWFHPGVNRCYCPSQEVAKRALFDGLDE 240

Qy    170 SNVKITGIPIRPQFEESMPV-GPIYKKYNLSPNKKVLLIMAGAHGV--LKNVKELCENLV 226
Db    241 SQVRVFGLPVRPSFARAVLVKDDLRKELEMDQDLRAVLLMGGGEGMGPVKETAKALEEFL 300

Qy    227 KDDQ-----VQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKP 281
Db    301 YDKENRKPIGQMVVICGRNKKLASALEAIDWK--IPVKVRGFETQMEKWMGACDCIITKA 358

Qy    282 GGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADE-D 340
Db    359 GPGTIAESLIRSLPIILNDYIPGQEKGNVPYVVENGAGVFTRSPKETARIVGEWFSTKTD 418

Qy    341 TLHRMKKNIKDLHLANSSEVILEDILKESE 370
Db    419 ELEQTSDNARKLAQPEAVFDIVKDIDELSE 448
```

RESULT 9
JC1275
phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) - Bacillus subtilis
N;Alternate names: murG protein; UPD-N-acetylglucosamine--N-acetylmuramyl-(pentapepti
C;Species: Bacillus subtilis
C;Date: 30-Sep-1993 #sequence_revision 06-Feb-1995 #text_change 20-Jun-2000
C;Accession: S25763; JC1275; S26499; G69662; S22212
R;Henriques, A.; de Lencastre, H.; Piggot, P.
Biochimie 74, 735-748, 1992
A;Title: A Bacillus subtilis morphogene cluster that includes spoVE is homologous to
A;Reference number: S25762; MUID:93003529
A;Accession: S25763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-363 <HEN>
A;Cross-references: EMBL:X64259; NID:g39994; PIDN:CAA45558.1; PID:g39995
R;Miyao, A.; Yoshimura, A.; Sato, T.; Yamamoto, T.; Theeragool, G.; Kobayashi, Y.
Gene 118, 147-148, 1992
A;Title: Sequence of the Bacillus subtilis homolog of the Escherichia coli cell-divis
A;Reference number: JC1275; MUID:92380484
A;Accession: JC1275
A;Molecule type: DNA
A;Residues: 1-180,'G',182-269,'T',271-363 <MIY>
A;Cross-references: GB:D10602; DDBJ:D90538; NID:g216299; PIDN:BAA01454.1; PID:g216300
R;Beall, B.; Lutkenhaus, J.
J. Bacteriol. 171, 6821-6834, 1989

A;Title: Nucleotide sequence and insertional inactivation of a Bacillus subtilis gene th
A;Reference number: S26499; MUID:90078133
A;Accession: S26499
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 216-363 <BEA>
A;Cross-references: EMBL:M31827; NID:g142831; PIDN:AAA83968.1; PID:g142832
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033
A;Accession: G69662
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-363 <KUN>
A;Cross-references: GB:Z99111; GB:AL009126; NID:g2633699; PIDN:CAB13395.1; PID:g2633893
A;Experimental source: strain 168
C;Genetics:
A;Gene: murG
C;Superfamily: murG protein
C;Keywords: glycosyltransferase; hexosyltransferase; peptidoglycan biosynthesis

Query Match 9.9%; Score 196; DB 2; Length 363;
Best Local Similarity 23.1%; Pred. No. 3.9e-06;
Matches 93; Conservative 76; Mismatches 153; Indels 80; Gaps 20;

```
Qy       6 RVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLY------------QESNPIVS- 51
           |: |     | ||:  |    ::  |   :|  |  ||              :|: |  |
Db       2 RIAISGGGTG-GHIYPALAFIKEVQR---RHPNVEFLYIGTENGLEKKIVERENIPFRSI 57

Qy      52 EVTQYLYLKSFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINT 111
           |:| :    ||   |   |                  |  | |:    : | :|| :| | |
Db      58 EITGFKRKLSFENVKTVMR-----------------FLKGVKKSKSYLAEFKPDAVIGT 99

Qy     112 FPMIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHEN------VDKYYVATDYVKEKLL-- 163
             :  |          |||              |||        :|:   : || :  :
Db     100 GGYVCGPVVYAAAKMGIPTI-----------VHEQNSLPGITNKF--LSKYVNKVAICF 145

Qy     164 -EIGTH-PS-NVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHG---VLKN 217
            |  :| ||  |  || | |     |:  |    :: || :|| :||  |: |   : :
Db     146 EEAKSHFPSEKVVFTGNP-RASEVVSIKTGRSLAEFKLSEDKKTVLIFGGSRGAAPINRA 204

Qy     218 VKELCENLVKDDQVQVVVVCGKNTALKESLSALEAEN-GDKLKVLGYVERIDELFRITDC 276
           | :: ::::|    ||: : |:    :: :: |:::   | :    :: :: |  :  |
Db     205 VIDM-QDVLKTRDYQVLYITGE-VHYEKVMNELKSKGAADNMVTKPFLHQMPEYLKAIDV 262

Qy     277 MITKPGGITLTEATAIGVPVILYKPVP----GQEKENANFFEDRGAAIVVNR----HEEI 328
           :: : |  |: | ||:|:| :|  | |       :: ||       ||||:        |::
Db     263 IVARAGAATIAEITALGIPSVLI-PSPYVTANHQEVNARSLGQHDAAIVLKETELSGEKL 321

Qy     329 LESVTSLLADEDTLHRMKKNIKDLHLANSSE---VILEDILK 367
           :|::  :: :| ||  | :  | | : :::     :||:: |
Db     322 IEALDRIVLNEQTLKEMSERTKSLGVPDAAARLYSVLEELKK 363
```

RESULT 10
C70401
phospho-N-acetylmuramoyl-pentapeptide transferase - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C;Accession: C70401
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666
A;Accession: C70401
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-344 <AQF>
A;Cross-references: GB:AE000727; NID:g2983623; PIDN:AAC07193.1; PID:g2983625; GB:AE00
A;Experimental source: strain VF5
C;Genetics:
A;Gene: murG
C;Superfamily: murG protein

Query Match 9.9%; Score 195; DB 2; Length 344;
Best Local Similarity 23.3%; Pred. No. 4.3e-06;
Matches 79; Conservative 57; Mismatches 133; Indels 70; Gaps 14;

```
Qy      41 NLYQESNPI------VSEVTQYLYLKSFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNK 94
             | ||:  |       |: ::|       |: ||   :|    |      |   :  |
Db      62 NFYQKLKAIWKFLKAQEEINEFL--------KEDYRALIFGG---YASLPLGINTVLRRK 110

Qy      95 RLGELVDEHQPDIIINTFPMIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVA 154
             |   : |       |: |        ::  :|: |||    |
Db     111 EL--FIHEQ------NSIPSKTNKILSKKAKKVLITFNYTKRF----------------- 145

Qy     155 TDYVKEKLLEIGTHPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHGV 214
                       |  |:: |:||| : :: :|   : |:: | |:|  :||  |: |
Db     146 -------------FPEGVRV-GLPIRKELKKKLPKKEVKKRFGLEPDKITVLIFGGSQGA 191

Qy     215 LKNVKELCENL--VKDDQVQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFR 272
           |  : ||  :|  |   : ||::: ||      |    ||   |:| :|: :   :  ::
Db     192 L-FLNELARDLKSVLPKEFQVILLTGK--IHYEKFKNLE---GEKFRVMPFSLDMGLIYS 245

Qy     273 ITDCMITKPGGITLTEATAIGVPVILYKPVP----GQEKENANFFEDRGAAIVVNRHEEI 328
           :|  |:: |  |: | :  ||| : : | |       :  ||   |  |  :|: : |
Db     246 ASDVAISRAGAGTINELSHFGVPSV-FVPYPYAVDDHQFYNAKEIEKLGGGLVLRQEEAK 304

Qy     329 LESVTSLLAD-EDTLHRMKKNIKDLHLANSSEVILEDIL 366
           : | | | :    | |  :|||      : | : |::|
Db     305 PDKVLSALKEIVKNLERYSENIKKFFAEGAEERMYEELL 343
```

RESULT 11
E83970
UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)pyrophos murG [imported] - Bac
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000
C;Accession: E83970
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A;Reference number: A83650; MUID:20263314
A;Accession: E83970
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-363 <STO>
A;Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BAB06284.1; GSPDB:G
A;Experimental source: strain C-125
C;Genetics:
A;Gene: murG
C;Superfamily: murG protein

Query Match 9.0%; Score 178; DB 2; Length 363;
Best Local Similarity 23.1%; Pred. No. 6.7e-05;
Matches 72; Conservative 62; Mismatches 124; Indels 54; Gaps 14;

```
Qy      87 IYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHE 146
           : |  | ||   |::| :||::| |   :  |         |||             :||
```

```
Db     75 VRFLRGTKRAKALLNEFKPDVVIGTGGYVCGPVVYAAAKLKIPTV----------IHE 122

Qy    147 N------VDKY---YVATDYVKEKLLEIGTHPSNVKITGIPIRPQF-----EESMPVGPI 192
Db    123 QNSVPGLTNKFLSRYVDRIAICFKEAEAFFPKNKVVFTGNPRASEVMSGNREEGL----- 177

Qy    193 YKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQV---QVVVVCG--KNTALKESL 247
Db    178 -RSLGIKPNKKTVLIVGGSRGA-RPINDAFMSILSDVKAKPYQFVYVTGTVHYERVQEQM 235

Qy    248 SAL-EAENGDKLKVLGYVERIDELFRITDCMITKPGGITLTEATAIGVPVILYKPVP--- 303
Db    236 KSIGQPEN---VIVQPFIHNMPDVLSAVDLIVARAGATTLAEITALGLPSILI-PSPYVT 291

Qy    304 -GQEKENANFFEDRGAAIVVNRHE----EILESVTSLLADEDTLHRMKKNIKDLHLANSS 358
Db    292 NNHQEKNAAALSKKDAAILRKESELTGDRLLEDIDDIMVTPGRLDAMKQAAKALGVPTAA 351

Qy    359 E---VILEDILK 367
Db    352 EKLHMLVKEVAK 363
```

RESULT 12
D64185
UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-ac
C;Species: Haemophilus influenzae
C;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 26-Aug-1999
C;Accession: D64185
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630
A;Accession: D64185
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-351 <TIGR>
A;Cross-references: GB:U32793; GB:L42023; NID:g1574683; PIDN:AAC22793.1; PID:g1574693; T
C;Genetics:
A;Gene: murG
C;Function:
A;Description: involved in murein or cell envelope biosynthesis
A;Pathway: peptidoglycan biosynthesis
C;Superfamily: murG protein
C;Keywords: cell division; cell wall; glycosyltransferase; hexosyltransferase; peptidogl

Query Match 8.7%; Score 171; DB 2; Length 351;
Best Local Similarity 23.9%; Pred. No. 0.00019;
Matches 95; Conservative 72; Mismatches 150; Indels 80; Gaps 22;

```
Qy      4 NKRVLILTANYGNGHV----QVAKTLYEQ---CVRLGFQHVTVSNLYQESNPIVSEVTQY 56
Db      3 NKKLLVMAGGTG-GHVFPAIAVAQTLQKQEWDICWLG----TKDRMEAQLVPKYGIPIRF 57

Qy     57 LYLKSFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIV 116
Db     58 IQISGLR-GK--------GIKALLN-APFAIFRAVLQAK--KIIQEEKPDAVLG------ 99

Qy    117 VPEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLL-EIGT-----HPS 170
Db    100 ------MGGYVSGPAGVAAKLCGVPIILHEQ-----NAIAGLTNKLLGKIATCVLQAFPT 148

Qy    171 ---NVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGA-------HGVLKNVKE 220
Db    149 AFPHAEVVGNPVREDLFE-MPNPDI--RFSDREEKLRVLVVGGSQGARVLNHTLPKVVAQ 205

Qy    221 LCENLVKDDQVQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITK 280
Db    206 LADKLEFRHQV------GKGAV--EEVSQLYGENLEQVKITEFIDNMAEAYAWADVVICR 257

Qy    281 PGGITLTEATAIGVPVILYKPVPGQEKE-----NANFFEDRGAAIVVNRHE---EILESV 332
Db    258 SGALTVCEIAAVGAAAIF---VPFQHKDRQQYLNAKYLSDVGAAKIIEQADLTPEILVNY 314

Qy    333 TSLLADEDTLHRMKKNIKDLHLANSSEVILEDILKES 369
Db    315 LKNLTRENLL-QMALKAKTMSMPNAAQRVAEVIKQYS 350
```

RESULT 13
F70195
UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 24-Nov-1999
C;Accession: F70195
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943
A;Accession: F70195
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-363 <KLE>
A;Cross-references: GB:AE001176; GB:AE000783; NID:g2688699; PIDN:AAC67113.1; PID:g268
A;Experimental source: strain B31
C;Superfamily: murG protein

Query Match 8.7%; Score 170.5; DB 2; Length 363;
Best Local Similarity 20.6%; Pred. No. 0.00022;
Matches 82; Conservative 84; Mismatches 171; Indels 61; Gaps 16;

```
Qy      3 TNKRVLILTANYGNGHV----QVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLY 58
Db      2 SNKKIIFFTGGGTGGHVFPGISIIQKLKEFDNEIEFFWIGKKNSIEEK--LIKEQDNIKF 59

Qy     59 LKSFSIGKQFYRLFYYGVDKIYNKRKFNIYFK--MGNKRLGELVDEHQPDIIINTFPMIV 116
Db     60 I-SIPCGK---------LRRYFSFKNFTDFFKVILGIIKSFYVLKKYKPQLIYATGGFVS 109

Qy    117 VPEYRRRTGRVIPTFNVMTDFCLHKIWVHE---------NVDKYYVATDYVKEKLLE-IG 166
Db    110 TPAI------------IASSLLKIKSITHEMDLDPGLATKINSKFANNIHISFKESEKYF 157

Qy    167 THPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHG--VLKNVKELCEN 224
Db    158 KNYKNIIYTGSPIRREFLNPDP--KIIKQLTQNTNKPIISILGGSLGANALNNL-ALC-- 212

Qy    225 LVKDDQVQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGI 284
Db    213 IKKDAEIYFIHQSGKN------LNDLSEKNYLRRQFFN-AEEMASIVKFSNLIISRAGAG 265

Qy    285 TLTEATAIGVPVILY---KPVPGQEKENANFFEDRGAAIVVNRHE----EILESVTSLLA 337
Db    266 AIKEFANAGACAILIPFKKGSRGDQIKNAKLLTNQNACIYIDEDEILNINILKIIKKTLK 325

Qy    338 DEDTLHRMKKNIKDLHLANSSEVILEDILKESEMMTAK 375
Db    326 DREKINSLKENIKKFNNKHSSTLIAKLLIKDIKETKSK 363
```

RESULT 14
E70156
lipopolysaccharide biosynthesis-related protein homolog - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 17-Mar-2000
C;Accession: E70156
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu

; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943
A;Accession: E70156
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-383 <KLE>
A;Cross-references: GB:AE001150; GB:AE000783; NID:g2688359; PIDN:AAC66815.1; PID:g268836
A;Experimental source: strain B31
C;Superfamily: probable hexosyltransferase ytxN

Query Match 8.3%; Score 163; DB 2; Length 383;
Best Local Similarity 22.7%; Pred. No. 0.00076;
Matches 80; Conservative 70; Mismatches 130; Indels 72; Gaps 21;

```
Qy      74 YGVDKIYNKRKFNIYFKMGNKR-LGELVDEHQPDII--INTFPMIVVPEYRRRTGR---- 126
           |    |   :| :      ||| : :::  ::||||   : | |        : |:
Db      52 YRCSSIQINKKLDAVIAFPNKRKISKIIQSYKPDIIHTHSEFSM-------GKIGKQIAL 104

Qy     127 -----VIPTFNVMTDFCLHKIWV-------------HENVDKYYV-ATDYVKEKLLEIGT 167
                ::| : | |: || : :              | |  |:::  :   ||:  ::
Db     105 KHNIPIVHTSHTMWDYYLHYLGIFKYFIKPDKMMRKHYNKIKHFIYPSSKAKERYFQLSN 164

Qy     168 HPSNVKI--TGIPIRPQFEESM---PVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELC 222
           : || ||   |:  |  | :::       | ||:|:    |: :|  |     ||:  |
Db     165 NSSNYKIIPNGVD-RKLFIKTLSKEKKDEILKKHNIKQTDKI-IIFVGRINKEKNINLLV 222

Qy     223 ENLVKDDQVQ----VVVVCGKNTALKESLSALEAENG--DKLKVLGYV--ERIDELFRIT 274
            :| ||  :|     ::: || :  || :      ::|   :: ::| :  | |   ::|:
Db     223 THL-KDLLMQNNNYKLILIGKGSEEKE-IKNFSIKHGLEKQILLIGTIPWEEIYYYYKIS 280

Qy     275 DCMITKPGG----ITLTEATAIGVPVIL-----YKPVPGQEKENANFFEDRGAAIVVNRH 325
           |   :        :|: ||   |:| ||     || |    ||  | |        :: ::
Db     281 DIFASLSKSEVYPMTVIEALTAGIPAILINDYIYKDV---IKEGINGF-------LIKKY 330

Qy     326 EEILESVTSLLADEDTLHRMKKNIKDLHLANSSEVILEDILKE--SEMMTAK 375
           | :   :  :: |:: | : |:| |  |    |        :|   ||::  |
Db     331 ENLSRYIDKVIKDDEILKKFKENAKK-HSTKFSSYFFTKKIKNYYSEIIARK 381
```

RESULT 15
E86823
peptidoglycan synthesis protein MurG [imported] - Lactococcus lactis subsp. lactis (stra
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001
C;Accession: E86823
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. in press, 2001
A;Title: The complete genome sequence of the lactic acid bacterium.
A;Reference number: A86625
A;Accession: E86823
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-357 <STO>
A;Cross-references: GB:AE005176; NID:g12724594; PIDN:AAK05687.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: murG
C;Superfamily: murG protein

Query Match 8.1%; Score 159; DB 2; Length 357;
Best Local Similarity 20.9%; Pred. No. 0.0013;
Matches 72; Conservative 51; Mismatches 129; Indels 92; Gaps 13;

```
Qy      54 TQYLYLKSFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFP 113
           | | : || |  |                              ::: | :||::: |
Db      73 TAYKFFKSVSDAK-----------------------------KIMKEFKPDVVLGTGG 101

Qy     114 MIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKL--LEIGTH--- 168
             :  |         |||             :||      :   :: :|:  : :| |
Db     102 YVAGPVVYAAAQLKIPTI-----------IHEGNSFPGITNRFLAKKVDRIAVGFHAAE 149

Qy     169 ----PSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCEN 224
                 |    ||    |: :|        :|:      :  ::|  |: | ||      |
Db     150 QYFPASKTTFTG---NPRAQEVADAAAQVEKF----EEPTVVIFGGSRGALKLNNAFIEA 202

Qy     225 L--VKDDQVQVVVVCGK--NTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITK 280
           |  :      : |    |:      ||: :  : :     | :  |:  : ||   :   : :
Db     203 LPELAQRSFKTVYASGEIYYDDYKETFN--QYKENSNLDIRPYINNMTELLAKSQLFLGR 260

Qy     281 PGGITLTEATAIGVPVILYKPVPG----QEKENANFFEDRGAAIVVNRHEEILESVTSLL 336
           |  |: | ||:|:| : | | |     |: :||  : |:||||::
Db     261 SGSTTIAEVTALGLPAV-YVPSPNVTADQQTKNAQEYVDQGAAIIIK------------ 306

Qy     337 ADEDTLHRMKKNIKDLHLANSSEVILEDILKESEMMTAKQKAKV 380
           |||        :    |  :   |||:  |  ||  |  || |
Db     307 -DED--------LTGQTLVEAISNILENNEKYQEMQAASLKAGV 341
```

Search completed: June 29, 2001, 09:00:19
Job time: 181 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 29, 2001, 08:59:08 ; Search time 16.33 Seconds
(without alignments)
801.322 Million cell updates/sec

Title: US-09-668-788-2
Perfect score: 1970
Sequence: 1 MNTNKRVLILTANYGNGHVQ.....EDILKESEMMTAKQKAKVLS 382

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	1970	100.0	382	1	YPFP_BACSU	P54166	bacillus su
2	196	9.9	363	1	MURG_BACSU	P37585	bacillus su
3	195	9.9	344	1	MURG_AQUAE	O67238	aquifex aeo
4	185.5	9.4	363	1	MURG_ENTFA	O07109	enterococcu
5	171	8.7	351	1	MURG_HAEIN	P45065	haemophilus
6	170.5	8.7	363	1	MURG_BORBU	O51708	borrelia bu
7	161.5	8.2	360	1	MURG_ENTHR	O07670	enterococcu
8	138.5	7.0	362	1	MURG_STRCU	Q9x4h4	streptomyce
9	137.5	7.0	364	1	MURG_STRCO	Q9zba5	streptomyce
10	137.5	7.0	407	1	MURG_MYCLE	O69552	mycobacteri
11	136	6.9	353	1	MURG_HELPY	O25770	helicobacte
12	135	6.9	484	1	YA62_METJA	Q58462	methanococc
13	134	6.8	353	1	MURG_HELPJ	Q9zk59	helicobacte
14	128.5	6.5	410	1	MURG_MYCTU	O06224	mycobacteri
15	123.5	6.3	480	1	NIFE_ANASP	Q44144	anabaena sp
16	123	6.2	354	1	MURG_ECOLI	P17443	escherichia
17	122.5	6.2	339	1	MURG_THEMA	Q9wy74	thermotoga
18	121.5	6.2	385	1	MURG_RICPR	Q9zdc0	rickettsia
19	117.5	6.0	380	1	YVYH_BACSU	P39131	bacillus su
20	116	5.9	820	1	SUS2_TULGE	Q41607	tulipa gesn
21	113.5	5.8	354	1	MURG_BUCAI	P57311	buchnera ap
22	113	5.7	371	1	MURG_SYNY3	P74657	synechocyst
23	112.5	5.7	1068	1	SPS_MAIZE	P31927	zea mays (m
24	110	5.6	529	1	UD16_RAT	P08430	rattus norv
25	106	5.4	1251	1	RBP2_PLAVB	Q00799	plasmodium
26	105.5	5.4	369	1	RFAB_ECOLI	P27127	escherichia
27	105.5	5.4	377	1	YPJH_BACSU	P42982	bacillus su
28	105.5	5.4	520	1	UD17_MOUSE	Q62452	mus musculu
29	105.5	5.4	1058	1	CARB_LACPL	P77886	lactobacill
30	105.5	5.4	1220	1	SLN1_YEAST	P39928	saccharomyc
31	105	5.3	1049	1	SPS_ORYSA	Q43802	oryza sativ
32	104	5.3	373	1	YC08_KLEPN	Q48454	klebsiella
33	104	5.3	1634	1	DPOL_METJA	Q58295	methanococc
34	102.5	5.2	487	1	UFO5_MANES	Q40287	manihot esc
35	102.5	5.2	531	1	UD16_MOUSE	Q64435	mus musculu
36	102	5.2	803	1	SUSY_ALNGL	P49034	alnus gluti
37	101	5.1	403	1	IF4A_LEIBR	Q25225	leishmania
38	100	5.1	530	1	UD18_RAT	Q64634	rattus norv
39	100	5.1	682	1	YE89_METJA	Q58884	methanococc
40	99	5.0	353	1	RFBU_SALTY	P26402	salmonella
41	99	5.0	376	1	WECB_ECOLI	P27828	escherichia
42	99	5.0	533	1	UD12_MOUSE	P70691	mus musculu
43	99	5.0	1045	1	SPS_BETVU	P49031	beta vulgar
44	99	5.0	1054	1	SPS1_CRAPL	O04932	craterostig
45	99	5.0	1875	1	MLP1_YEAST	Q02455	saccharomyc

ALIGNMENTS

RESULT 1
YPFP_BACSU
ID YPFP_BACSU STANDARD; PRT; 382 AA.
AC P54166;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 43.6 KDA PROTEIN IN CPSD-METB INTERGENIC REGION.
GN YPFP.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96349105; PubMed=8760912;
RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,
RA Serror P.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the serA and kdg loci cloned in a yeast artificial chromosome.";
RL Microbiology 142:2005-2016(1996).
CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; L77246; AAA96624.1; -.
DR EMBL; Z99115; CAB14110.1; -.
DR SubtiList; BG11611; ypfP.
DR InterPro; IPR001296; -.
DR Pfam; PF00534; Glycos_transf_1; 1.
KW Hypothetical protein; Transferase; Glycosyltransferase.
SQ SEQUENCE 382 AA; 43561 MW; 8A8B98593DAC6041 CRC64;

Query Match 100.0%; Score 1970; DB 1; Length 382;
Best Local Similarity 100.0%; Pred. No. 9.8e-129;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MNTNKRVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MNTNKRVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLK 60

Qy     61 SFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 SFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEY 120

Qy    121 RRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKITGIPIR 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 RRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKITGIPIR 180
```

```
Qy     181 PQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVVVCGKN 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 PQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVVVCGKN 240

Qy     241 TALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGITLTEATAIGVPVILYK 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGITLTEATAIGVPVILYK 300

Qy     301 PVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDTLHRMKKNIKDLHLANSSEV 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 PVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDTLHRMKKNIKDLHLANSSEV 360

Qy     361 ILEDILKESEMMTAKQKAKVLS 382
           ||||||||||||||||||||||
Db     361 ILEDILKESEMMTAKQKAKVLS 382


RESULT  2
MURG_BACSU
ID   MURG_BACSU     STANDARD;      PRT;   363 AA.
AC   P37585; P18578; Q59247;
DT   01-NOV-1990 (Rel. 16, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92380484; PubMed=1387377;
RA   Miyao A., Yoshimura A., Sato T., Yamamoto T., Theeragool G.,
RA   Kobayashi Y.;
RT   "Sequence of the Bacillus subtilis homolog of the Escherichia coli
RT   cell-division gene murG.";
RL   Gene 118:147-148(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=93003529; PubMed=1391053;
RA   Henriques A.O., de Lencastre H., Piggot P.J.;
RT   "A Bacillus subtilis morphogene cluster that includes spoVE is
RT   homologous to the mra region of Escherichia coli.";
RL   Biochimie 74:735-748(1992).
RN   [3]
RP   SEQUENCE OF 216-363 FROM N.A.
RX   MEDLINE=90078133; PubMed=2556375;
RA   Beall B., Lutkenhaus J.;
RT   "Nucleotide sequence and insertional inactivation of a Bacillus
RT   subtilis gene that affects cell division, sporulation, and
RT   temperature sensitivity.";
RL   J. Bacteriol. 171:6821-6834(1989).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D10602; BAA01454.1; -.
DR   EMBL; X64259; CAA45558.1; -.
DR   EMBL; M31827; AAA83968.1; -.
DR   EMBL; Z99111; CAB13395.1; -.
DR   PIR; JC1275; JC1275.
DR   PIR; S26499; S26499.
DR   SubtiList; BG10227; murG.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
FT   CONFLICT    181    181       K -> G (IN REF. 1).
FT   CONFLICT    270    270       A -> T (IN REF. 1).
SQ   SEQUENCE   363 AA;  39977 MW;  185B20688C732489 CRC64;


  Query Match             9.9%;  Score 196;  DB 1;  Length 363;
  Best Local Similarity   23.1%;  Pred. No. 1.4e-06;
  Matches   93;  Conservative   76;  Mismatches  153;  Indels   80;  Gaps   20;

Qy       6 RVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLY------------QESNPIVS- 51
           |: |     | ||:  |    ::  |   :|  |  ||             :|: |  |
Db       2 RIAISGGGTG-GHIYPALAFIKEVQR---RHPNVEFLYIGTENGLEKKIVERENIPFRSI 57

Qy      52 EVTQYLYLKSFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINT 111
           |:| :    ||   |  |                  |  | |:    : | :|| :| |
Db      58 EITGFKRKLSFENVKTVMR-----------------FLKGVKKSKSYLAEFKPDAVIGT 99

Qy     112 FPMIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHEN------VDKYYVATDYVKEKLL-- 163
             :  |         |||             |||        :|:   : || :  :
Db     100 GGYVCGPVVYAAAKMGIPTI-----------VHEQNSLPGITNKF--LSKYVNKVAICF 145

Qy     164 -EIGTH-PS-NVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHG---VLKN 217
            |  :| ||  |  || | |     |:  |    :: || :|| :||  |: |   : :
Db     146 EEAKSHFPSEKVVFTGNP-RASEVVSIKTGRSLAEFKLSEDKKTVLIFGGSRGAAPINRA 204

Qy     218 VKELCENLVKDDQVQVVVVCGKNTALKESLSALEAEN-GDKLKVLGYVERIDELFRITDC 276
           | :: ::::|    ||: : |:    :: :: |:::   | :    :: :: |  :  |
Db     205 VIDM-QDVLKTRDYQVLYITGE-VHYEKVMNELKSKGAADNMVTKPFLHQMPEYLKAIDV 262

Qy     277 MITKPGGITLTEATAIGVPVILYKPVP----GQEKENANFFEDRGAAIVVNR----HEEI 328
           :: : |  |: | ||:|:| :|  | |      :: ||       ||||:       |::
Db     263 IVARAGAATIAEITALGIPSVLI-PSPYVTANHQEVNARSLGQHDAAIVLKETELSGEKL 321

Qy     329 LESVTSLLADEDTLHRMKKNIKDLHLANSSE---VILEDILK 367
           :|::  :: :| ||  | :  | | : :::     :||:: |
Db     322 IEALDRIVLNEQTLKEMSERTKSLGVPDAAARLYSVLEELKK 363


RESULT  3
MURG_AQUAE
ID   MURG_AQUAE     STANDARD;      PRT;   344 AA.
AC   O67238;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG OR AQ_1177.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
```

```
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000727; AAC07193.1; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   344 AA;  39325 MW;  5A4B3322B4AC650C CRC64;


  Query Match             9.9%;  Score 195;  DB 1;  Length 344;
  Best Local Similarity   23.3%;  Pred. No. 1.5e-06;
  Matches   79;  Conservative   57;  Mismatches  133;  Indels   70;  Gaps   14;

Qy      41 NLYQESNPI------VSEVTQYLYLKSFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNK 94
           | ||:   |       |: ::|       |: ||   :|   |      |   :  |
Db      62 NFYQKLKAIWKFLKAQEEINEFL--------KEDYRALIFGG---YASLPLGINTVLRRK 110

Qy      95 RLGELVDEHQPDIIINTFPMIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVA 154
             |   : |      |: |       ::  :|: |||    |
Db     111 EL--FIHEQ------NSIPSKTNKILSKKAKKVLITFNYTKRF----------------- 145

Qy     155 TDYVKEKLLEIGTHPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHGV 214
                       |  |:: |:||| : :: :|   : |:: |.|:|  :||  |: |
Db     146 ------------FPEGVRV-GLPIRKELKKKLPKKEVKKRFGLEPDKITVLIFGGSQGA 191

Qy     215 LKNVKELCENL--VKDDQVQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFR 272
           |  : ||  :|  |   : ||::: ||      |    ||   |:| :|: :   :  ::
Db     192 L-FLNELARDLKSVLPKEFQVILLTGK--IHYEKFKNLE---GEKFRVMPFSLDMGLIYS 245

Qy     273 ITDCMITKPGGITLTEATAIGVPVILYKPVP----GQEKENANFFEDRGAAIVVNRHEEI 328
           :|  |:: |  |: | :  ||| : : | |       :  ||   |  |  :|: : |
Db     246 ASDVAISRAGAGTINELSHFGVPSV-FVPYPYAVDDHQFYNAKEIEKLGGGLVLRQEEAK 304

Qy     329 LESVTSLLAD-EDTLHRMKKNIKDLHLANSSEVILEDIL 366
           : | | | :    | |  :|||      : | : |::|
Db     305 PDKVLSALKEIVKNLERYSENIKKFFAEGAEERMYEELL 343


RESULT   4
MURG_ENTFA
ID   MURG_ENTFA     STANDARD;      PRT;   363 AA.
AC   O07109;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
OX   NCBI_TaxID=1351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A24836;
RX   MEDLINE=97431524; PubMed=9287029;
RA   Pucci M.J., Thanassi J.A., Discotto L.F., Kessler R.E.,
RA   Dougherty T.J.;
RT   "Identification and characterization of cell wall-cell division gene
RT   clusters in pathogenic gram-positive cocci.";
RL   J. Bacteriol. 179:5632-5635(1997).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U94707; AAC45636.1; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   363 AA;  39916 MW;  5178BCC95264BE69 CRC64;


  Query Match             9.4%;  Score 185.5;  DB 1;  Length 363;
  Best Local Similarity   22.3%;  Pred. No. 7.2e-06;
  Matches   91;  Conservative   76;  Mismatches  144;  Indels   97;  Gaps   19;

Qy       8 LILTANYGNGHVQVAKTLYE----QCVRLGFQHV-TVSNLYQESNPIVSEVTQYLYLKSF 62
           :::|     ||:  | :  |     :     | :| | : |  :  |      : : :: |
Db       3 ILVTGGGTGGHIYPALSFVEHVKKEAPATEFLYVGTENGLESQIVPKAKIPFKTIKIQGF 62

Qy      63 --SIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEY 120
             |:  | ::  |  :  | || |              ::: | |||::| |   :
Db      63 KRSLSPQNFKTIYLFLTSI-NKAK------------KIIREFQPDVVIGTGGYV----- 103

Qy     121 RRRTGRVIPTFNVMTDFCLHKI----WVHEN-------------VDKYY-----VATDY 157
             :| |:        :  |::     :||                 |||       ||: :
Db     104 ---SGAVV--------YAAHQLKIPTIIHEQNSIPGMTNKFLSRYVDKIAICFPDVASFF 152

Qy     158 VKEKLLEIGTHPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHGVLKN 217
           | ||| :           :|| | | |    ::    |  :: | | || :::  |: | ||
Db     153 PKEKTI----------LTGNP-RGQEVVTVEKSAILSEFGLDPAKKTVVLFGGSRGALK- 200

Qy     218 VKELCEN---LVKDDQVQVVVVCGKN--TALKESLSALEAENGDKLKVLGYVERIDELFR 272
           : :  |    | :: : ||:    |:      |:|||  |  :     : |  |:::: |:
Db     201 INQAFEQAFPLFEEREYQVLYASGERYYQELQESLK-LSEKKLTNISVQPYIDKMVEVMA 259

Qy     273 ITDCMITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESV 332
           || |:  : |  :: | ||:|:| ||   :|     | :  ::  : : |    | | ::
Db     260 NTDLMVGRAGATSIAEFTALGLPAIL---IPSPYVTNDHQTKNAQSLVKVGAVEMIPDA- 315

Qy     333 TSLLADEDTLHRMKKNIKDLHLANSSEVILEDILKESEMMTAKQKAKV 380
                 | |  |:   | |: | |          |  :| || :  ::
Db     316 ------ELTGARLVAAIDDILLNNE---------KRQQMATASKGERI 348


RESULT   5
MURG_HAEIN
ID   MURG_HAEIN     STANDARD;      PRT;   351 AA.
AC   P45065;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG OR HI1138.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U32793; AAC22793.1; -.
DR   TIGR; HI1138; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Inner membrane; Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   351 AA;  38307 MW;  41067F379BB30B27 CRC64;
```

```
  Query Match          8.7%;  Score 171;  DB 1;  Length 351;
  Best Local Similarity  23.9%;  Pred. No. 6.8e-05;
  Matches   95;  Conservative   72;  Mismatches  150;  Indels   80;  Gaps   22;

Qy       4 NKRVLILTANYGNGHV----QVAKTLYEQ---CVRLGFQHVTVSNLYQESNPIVSEVTQY 56
           ||:::|::    | |||     ||:|| :|      ||    |   :  :  |      ::
Db       3 NKKLLVMAGGTG-GHVFPAIAVAQTLQKQEWDICWLG----TKDRMEAQLVPKYGIPIRF 57

Qy      57 LYLKSFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIV 116
           : :    ||         |:  : |   | |:  :   :  ::: | :|| ::
Db      58 IQISGLR-GK--------GIKALLN-APFAIFRAVLQAK--KIIQEEKPDAVLG------ 99

Qy     117 VPEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLL-EIGT-----HPS 170
             | |     |    |   | :||       |   :  ||| :| |      |:
Db     100 ------MGGYVSGPAGVAAKLCGVPIILHEQ-----NAIAGLTNKLLGKIATCVLQAFPT 148

Qy     171 ---NVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGA-------HGVLKNVKE 220
             : :: | |:|    | ||   |  :::    |  :|:: |:       | : | | :
Db     149 AFPHAEVVGNPVREDLFE-MPNPDI--RFSDREEKLRVLVVGGSQGARVLNHTLPKVVAQ 205

Qy     221 LCENLVKDDQVQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITK 280
           | : |    ||      ||     | :| |  || :::|:  ::: : | :   | :| :
Db     206 LADKLEFRHQV------GKGAV--EEVSQLYGENLEQVKITEFIDNMAEAYAWADVVICR 257

Qy     281 PGGITLTEATAIGVPVILYKPVPGQEKE-----NANFFEDRGAAIVVNRHE---EILESV 332
             | :|: |  |:|   |    || | |:     || :  | ||| :: : :   ||| :
Db     258 SGALTVCEIAAVGAAAIF---VPFQHKDRQQYLNAKYLSDVGAAKIIEQADLTPEILVNY 314

Qy     333 TSLLADEDTLHRMKKNIKDLHLANSSEVILEDILKES 369
             |  |: | :|    | : : |::: : | | : |
Db     315 LKNLTRENLL-QMALKAKTMSMPNAAQRVAEVIKQYS 350

RESULT   6
```

```
MURG_BORBU
ID   MURG_BORBU     STANDARD;      PRT;   363 AA.
AC   O51708;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG OR BB0767.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE001176; AAC67113.1; -.
DR   TIGR; BB0767; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   363 AA;  41118 MW;  1BAFA347384DB235 CRC64;
```

```
  Query Match          8.7%;  Score 170.5;  DB 1;  Length 363;
  Best Local Similarity  20.6%;  Pred. No. 7.7e-05;
  Matches   82;  Conservative   84;  Mismatches  171;  Indels   61;  Gaps   16;

Qy       3 TNKRVLILTANYGNGHV----QVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLY 58
           :||:::  |     |||     : : | |    : |   |  :|   :: |      :
Db       2 SNKKIIFFTGGGTGGHVFPGISIIQKLKEFDNEIEFFWIGKKNSIEEK--LIKEQDNIKF 59

Qy      59 LKSFSIGKQFYRLFYYGVDKIYNKRKFNIYFK--MGNKRLGELVDEHQPDIIINTFPMIV 116
           : |   ||         : : :: : |  :||  :|  :   :: :::| :|  |   :
Db      60 I-SIPCGK---------LRRYFSFKNFTDFFKVILGIIKSFYVLKKYKPQLIYATGGFVS 109

Qy     117 VPEYRRRTGRVIPTFNVMTDFCLHKIWVHE---------NVDKYYVATDYVKEKLLE-IG 166
             |                : :     |   ||          ::  :    ::  |  |
Db     110 TPAI------------IASSLLKIKSITHEMDLDPGLATKINSKFANNIHISFKESEKYF 157

Qy     167 THPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHG--VLKNVKELCEN 224
             :  |:  || ||| :|    |   | |:   : || :: |: |: |    | |:  ||
Db     158 KNYKNIIYTGSPIRREFLNPDP--KIIKQLTQNTNKPIISILGGSLGANALNNL-ALC-- 212

Qy     225 LVKDDQVQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGI 284
```

```
          : || ::  :   |||      |: |  :|  : :     | :  : : :: :|:: |
Db    213 IKKDAEIYFIHQSGKN------LNDLSEKNYLRRQFFN-AEEMASIVKFSNLIISRAGAG 265

Qy    285 TLTEATAIGVPVILY---KPVPGQEKENANFFEDRGAAIVVNRHE----EILESVTSLLA 337
          : |    |   ||    |   | : :||    :: | | ::  |     ||: :   |
Db    266 AIKEFANAGACAILIPFKKGSRGDQIKNAKLLTNQNACIYIDEDEILNINILKIIKKTLK 325

Qy    338 DEDTLHRMKKNIKDLHLANSSEVILEDILKESEMMTAK 375
          | : :: :|:|||  :  :|| :| : ::|: :   :|
Db    326 DREKINSLKENIKKFNNKHSSTLIAKLLIKDIKETKSK 363


RESULT   7
MURG_ENTHR
ID   MURG_ENTHR     STANDARD;      PRT;   360 AA.
AC   O07670;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG.
OS   Enterococcus hirae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
OX   NCBI_TaxID=1354;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S185;
RX   MEDLINE=99449055; PubMed=10520745;
RA   Duez C., Thamm I., Sapunaric F., Coyette J., Ghuysen J.-M.;
RT   "The division and cell wall gene cluster of Enterococcus hirae S185.";
RL   DNA Seq. 9:149-161(1998).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y13922; CAA74235.1; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   360 AA;  39591 MW;  444D6F7CFEC1FDA1 CRC64;


  Query Match             8.2%;  Score 161.5;  DB 1;  Length 360;
  Best Local Similarity   20.6%;  Pred. No. 0.00032;
  Matches   84;  Conservative   69;  Mismatches  149;  Indels  105;  Gaps   17;

Qy      8 LILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLKSF----- 62
          :::|     ||:  |         | | :   :   :| |      |:::|: :
Db      3 ILVTGGGTGGHIYPA---------LAFVNYVKT---KEPN------TEFMYVGAQRGLEN 44

Qy     63 ----SIGKQFYRLFYYGVDK---IYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMI 115
               |  |  |   |  :   ::| :   :: |   :   ::: | :||::| |   :
Db     45 KIVPETGMPFRTLEIQGFQRKLSLHNLKTIQLFLK-SIREAKKILKEFKPDVVIGTGGYV 103

Qy    116 VVPEYRRRTGRVIPTFNVMTDFCLHKIWVHEN----------VDKYYVATD----- 156
              :  |||           :||                  ||:  :: :
Db    104 SGAVVYAASKLAIPTI----------IHEQNSVPGITNKFLSRYVDRIALSFEDAAPF 151

Qy    157 YVKEKLLEIGTHPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHGVLK 216
          :  ||   || :|          | |    |    |   | | | || :||  |: | ||
Db    152 FPAEKSSLIG-NP----------RAQEVADMDKSKILATYGLDPEKKTVLIFGGSQGALK 200

Qy    217 NVKELCENLVKDDQVQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDC 276
          : : | |:  ||   |:          |:  : : |     : :  |: :: |:   :|
Db    201 INQAVTEFLMSFDQEYQVLYASGERYYKDIQTKVPA--CANVSIQPYINKMAEVMASSDL 258

Qy    277 MITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLL 336
          :: : |  :: | ||:|:| ||  | |                 | | |:   ::  ||:
Db    259 LVGRAGATSIAELTALGLPAILI-PSP---------------YVTNDHQ--TKNAMSLV 299

Qy    337 ADEDTLHRMKKNIKDLHLANSS------EVILEDILKESEMMTAKQK 377
              :  | |||  |   |       |:  | |::   : :||:
Db    300 K-----NNAAKMIKDDELDGRSLKQAIEEIMTNDQLQKQMSLASKQQ 341


RESULT   8
MURG_STRCU
ID   MURG_STRCU     STANDARD;      PRT;   362 AA.
AC   Q9X4H4;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG.
OS   Streptomyces collinus.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=42684;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 40733;
RA   Mikulik K., Zhulanova E.;
RT   "Characterization of dcw cluster from Streptomycetes.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF110367; AAD26629.1; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   362 AA;  38847 MW;  0BBA131F2E5D5122 CRC64;


  Query Match             7.0%;  Score 138.5;  DB 1;  Length 362;
  Best Local Similarity   22.1%;  Pred. No. 0.012;
  Matches   46;  Conservative   47;  Mismatches   88;  Indels   27;  Gaps    8;

Qy    168 HPSNVKITGI---PIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHGV--LKNVKELC 222
          ||    |  :    :||:            :: | ||   ||:  |: |   |  | :
Db    158 HPVRRSIATLDRAAVRPEAR---------ARFGLDPNLPTLLVSGGSQGARRLNEVVQQV 208

Qy    223 ENLVKDDQVQVVVVCGKNTALKESLSALEAENGDKLKV-LGYVERIDELFRITDCMITKP 281
          ::   :|::   |     |  |  ::   |    : : |::|:|  :   | |: :
Db    209 APWLQQAGIQILHAVGP----KNELPQVQQMPGMPPYIPVSYLDRMDLAYAAADMMLCRA 264
```

```
Qy     282 GGITLTEATAIGVPVILYKPVP---GQEKENANFFEDRGAAIVVN----RHEEILESVTS 334
           | :|: | :|:|:|   | :|   |::: ||      |  ::|:      | : ::|
Db     265 GAMTVAELSAVGLPA-AYVPLPIGNGEQRLNAQPVVKAGGGLLVDDAELTPEWVQQTVLP 323

Qy     335 LLADEDTLHRMKKNIKDLHLANSSEVIL 362
           :|||   |: | :   :    :: |:::
Db     324 VLADPHRLYEMSRAAGEFGRRDADELLV 351


RESULT  9
MURG_STRCO
ID   MURG_STRCO     STANDARD;      PRT;   364 AA.
AC   Q9ZBA5;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG OR SC4A10.17C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Kuennen R.A., Stadelmaier B.T., McCormick J.R.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Saunders D.C., Harris D., James K.D., Parkhill J., Barrell B.G.,
RA   Rajandream M.A.;
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U10879; AAD10537.1; -.
DR   EMBL; AL109663; CAB51993.1; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   364 AA;  38842 MW;  D600C5F94E9202E2 CRC64;


  Query Match             7.0%;  Score 137.5;  DB 1;  Length 364;
  Best Local Similarity   19.9%;  Pred. No. 0.015;
  Matches   59;  Conservative   64;  Mismatches  119;  Indels   55;  Gaps   14;

Qy      94 KRLGELVDEHQPDIIINTFPMIVVPEY--RRRTGRVIPTFNVMTDFCLHKIWVHE----- 146
           |   ::::  : | :      : :| |  :| |  :|           | |||
Db      82 KATEQILERTKADAVAGFGGYVALPAYLAAKRLG--VP-----------IVVHEANARP 127

Qy     147 ----NVDKYYVATDYVKEKLLEIGTHPS---NVKITGIPIR---PQFEESMPVGPIYKKY 196
               :  | |       : : |  |   | :  |||:|      : :          :
Db     128 GLANKIGSRYAAQ-------VAVSTPDSKLRNSRYIGIPLRRSIATLDRAAARPEARAMF 180

Qy     197 NLSPNKKVLLIMAGAHGVLKNVKELCEN---VKDDQVQVVVVCGKNTALKESLSALEAE 253
           | ||   ||:  |: |  : : |: : :   ::   :|::   |     |  |  :
Db     181 GLDPNLPTLLVTGGSQGA-RRLNEVIQQVAPWLQQAGIQILHAVGP----KNELPQVHQM 235

Qy     254 NGDKLKV-LGYVERIDELFRITDCMITKPGGITLTEATAIGVPVILYKPVP---GQEKEN 309
           |    : : |::|:|  :   | |: : | :|: | :|:|:|   | |:|   |::: |
Db     236 PGMPPYIPVSYLDRMDLAYAAADMMLCRAGAMTVAELSAVGLPA-AYVPLPIGNGEQRLN 294

Qy     310 ANFFEDRGAAIVVN----RHEEILESVTSLLADEDTLHRMKKNIKDLHLANSSEVIL 362
           |      |  ::|:      | : ::|  :|||   |: | :   :    :: ::::
Db     295 AQPVVKAGGGLLVDDAELTPEWLQQNVLPVLADPHRLYEMSRAAAEFGRRDADDLLV 351


RESULT  10
MURG_MYCLE
ID   MURG_MYCLE     STANDARD;      PRT;   407 AA.
AC   O69552;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG OR MLCB268.02C.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brown D., Churcher C.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL022602; CAA18668.1; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   407 AA;  42422 MW;  B757B73C092C53F9 CRC64;


  Query Match             7.0%;  Score 137.5;  DB 1;  Length 407;
  Best Local Similarity   22.7%;  Pred. No. 0.017;
  Matches   68;  Conservative   49;  Mismatches  112;  Indels   71;  Gaps   15;

Qy      98 ELVDEHQPDIIINTFPMIVVPEYRRRTG-----RVIPTFNVMTDFCLHKIWVHENVDKYY 152
           |:|: |   :::     : :| |     |      | ||           : |||    :
Db     117 EVVEAH---VVVGFGGYVALPAYLAARGIPRVRRRIP-----------VVVHEANARAG 161

Qy     153 VATDY---VKEKLLEI--GTHPSNVKITGIPI--------RPQFEESMPVGPIYKKYNLS 199
           :|        |::|    |:     :: |:||        ||           | :  :
Db     162 IANRVGVRTAERVLSAVPGSGLRGAEVVGVPIHATITTLNRPALR-----ADARKHFGFT 216

Qy     200 PNKKVLLIMAGAHGVL---KNVKELCENLVKDDQVQVVVVCGKNTALKESLSALEAENGD 256
           : :|||:  |: | :   : |     |:|     | |:   |     || :|     | |:
Db     217 DDARVLLVFGGSQGAVSLNRAVAGAAEDLAASG-VAVLHAYG----LKNTLELRTPEYGE 271

Qy     257 KLKV-LGYVERIDELFRITDCMITKPGGITLTEATAIGVPVILYKPVP---GQEKENANF 312
           | : |::|:|  :   | :| : | :|: | :|:|:| | | | |   |::: ||
Db     272 PPYVAVPYLDRMDLAYAAADLVICRSGAMTVAEVSAVGLPAI-YVPFPIGNGEQRLNALP 330
```

```
Qy    313 FEDRGAAIVVNRHE----EILESVTSLLAD----------------EDTLHRMKKNIKDL 352
          : |  :||   :     :   |  | :|                  |  | : |   ||
Db    331 VVNAGGGLVVADADLTPGLVARQVVRLFSDPAQLAAMTAAAARVGHRDAAHHVAKVALDL 390


RESULT 11
MURG_HELPY
ID   MURG_HELPY     STANDARD;      PRT;   353 AA.
AC   O25770;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG OR HP1155.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000621; AAD08196.1; -.
DR   TIGR; HP1155; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   353 AA;  39474 MW;  6D136CD92A180513 CRC64;


  Query Match             6.9%;  Score 136;  DB 1;  Length 353;
  Best Local Similarity   19.5%;  Pred. No. 0.018;
  Matches   78;  Conservative   76;  Mismatches  150;  Indels   96;  Gaps   20;

Qy     10 LTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQ------ESNPIVSEVTQYLYLKSFS 63
          ||     ||: :|| |  :  : | : : : : |       |::|: ||  :| :
Db      5 LTGGGTGGHLSIAKALAIELEKQGIEAIYLGSTYGQDKEWFENSPLFSE--RYFFNTQGV 62

Qy     64 IGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEYRRR 123
          : | |::     :  :: : |     |       |:: :||    |:       |
Db     63 VNKSFFK----KIGSLFLQAKAAFKAK-------EILKKHQITHTISVGGFSAGPASFAS 111

Qy    124 TGRVIPTFNVMTDFCLHKIWVHE------NVDKYY------VATDYV-KEKLLEIGTHPS 170
                ||           :::||      ::::|       | : |  |:|    | |
Db    112 LLNKIP-----------LYIHEQNAIKGSLNRYLSPKAKAVFSSYAFKDK----GNH-- 153

Qy    171 NVKITGIPIRPQFEESMPVGPIYKKYNLSPNK---KVLLIMAGAHG-------VLKNVKE 220
          :|  |::  |            :: :  :   | :| : |: |         | |  :
Db    154 --VLTSYPVQNAF-----------FDFARTRTEIKHILFLGGSQGAKAINEFALLNAPK 199

Qy    221 LCENLVKDDQVQVVVVCGKNTALKESLSALEAENG--DKLKVLGYVERIDELFRITDCMI 278
          | :  :|     :  :|| |:   | :     | |  ||::::  :   | |:    | :
Db    200 LTKQGIK-----ITHICGPNS--YEQVRFFYQELGLLDKIELFAFHNNITEIMHRADLCV 252

Qy    279 TKPGGITLTEATAIGVPVILYKPVP----GQEKENANFFEDRGAAIVVNRHE----EILE 330
          :: |  :: |  | |:| | : | |      :  |   ||      || ::|    :: |
Db    253 SRAGASSVWELCANGLPTI-FIPYPFASNNHQYYNVLEFEKENLCYVVPQNELLPKKLFE 311

Qy    331 SVTSLLADED----TLHRMKKNIKDLHLANSSEVILEDIL 366
          :  |   :|     |  :   ::     : :: |:| ||
Db    312 VIRKLNQKDDQGNKNLTTISNQLQQKIAKDGAKTIIETIL 351


RESULT 12
YA62_METJA
ID   YA62_METJA     STANDARD;      PRT;   484 AA.
AC   Q58462;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   HYPOTHETICAL PROTEIN MJ1062.
GN   MJ1062.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: STRONG, TO B.SUBTILIS SPSG.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67549; AAB99065.1; -.
DR   TIGR; MJ1062; -.
DR   InterPro; IPR000182; -.
DR   Pfam; PF00583; Acetyltransf; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   484 AA;  56693 MW;  E1764675A97CFCB2 CRC64;


  Query Match             6.9%;  Score 135;  DB 1;  Length 484;
  Best Local Similarity   20.5%;  Pred. No. 0.032;
  Matches   78;  Conservative   63;  Mismatches  114;  Indels  126;  Gaps   19;

Qy     75 GVDKIYNK-------RKFN----IYFKMGNKRLGELVDEH------------------- 103
          |:  :|         ||||    |:|   :: : : ::|:
Db     14 GMGHVYRTLSLANELRKFNVNEIIFFTKSDEDVIKKIEENGFKVIKCSDNNDILKNIKNI 73
```

```
Qy     104 QPDIIINTFPMIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLL 163
           :||::|                  |    :  ||       :|:       :  |: :
Db      74 KPDVVI-----------------IDDLGIEEDFA-------KNI------RELCKKLIF 102

Qy     164 EIGTHPSNVKITGIPI----------RPQFEESMPV----GPIY----------KKYNLS 199
             :||: |   | :           |  |:|        || |           ||  ||
Db     103 FDNPNPSSNKYADIVVNAIVGSELKNRKYFDEENKTLYFYGPKYLILRNEFYKVKKEMLS 162

Qy     200 -----PNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVVVCGKNTALKESLSAL--EA 252
                | :|:   |:       | | | | ||    : || |     :: |: |
Db     163 RSKNKETKNILIAFGGSDPSNLTCKVLEELLSKDRDFNINVVLGPKFQYEDELNNLLKRY 222

Qy     253 ENGDKLKVLGYVERIDELFRITDCMITKPGGITLTEATAIGVP-VILYKPVPGQE----- 306
             ||:|:   :: : || :  | :|| | |:|: ||  :|:| |:||:    :|
Db     223 SKSDKIKIYKNIDNMAELMKDNDLIITSP-GMTMFEALFLGIPVVVLYQNELQRECYDDY 281

Qy     307 -----KENAN------FFEDRGAAIVVNRHE-EILESVTSLLADEDTLHRMKKNIKDLHL 354
                | : |      | :     : : : : ||:|::|:       ::  ||      :
Db     282 LKKISKTHLNPLKEGYFIDAEHTDLHIGKGKFEIIEAITN-------IYNCKK------I 328

Qy     355 ANSSEVILEDIL-KESEMMTA 374
             |::|:  |   : |:: |
Db     329 GEDSKIIIRQITDNDLELLMA 349
```

```
RESULT 13
MURG_HELPJ
ID   MURG_HELPJ     STANDARD;      PRT;   353 AA.
AC   Q9ZK59;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG OR JHP1082.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic-sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001535; AAD06652.1; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   353 AA;  39506 MW;  1773AD7E5C214EF9 CRC64;
```

```
  Query Match              6.8%;  Score 134;  DB 1;  Length 353;
  Best Local Similarity   20.0%;  Pred. No. 0.024;
  Matches   80;  Conservative   71;  Mismatches  152;  Indels   98;  Gaps   20;

Qy      10 LTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQ------ESNPIVSEVTQYLYLKSFS 63
           ||      ||: :|| |  :  : | : : : : |       |::|: ||  :| :
Db       5 LTGGGTGGHLSIAKALAIELEKQGIEAIYLGSTYGQDKEWFENSPLFSE--RYFFNTQGV 62

Qy      64 IGKQFYR----LFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVVPE 119
           : | |::    ||      :  |  |         :  |:: :||    |:       |
Db      63 VNKSFFKKIRSLF------LQAKAAF---------KAKEILKKHQITHTISVGGFSAGPA 107

Qy     120 YRRRTGRVIPTFNVMTDFCLHKIWVHE------NVDKYY------VATDYV-KEKLLEIG 166
                  ||             :::||      ::::| .      | : |  |:|    |
Db     108 SFASLLNKIP-----------LYIHEQNAIKGSLNRYLSPKAKAVFSSYAFKDK----G 151

Qy     167 THPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHG-------VLKNVK 219
           |     :|  |::  |          :          | :| : |: |         | |
Db     152 NH----VLTSYPVQNAF---------FDHARTRTEIKHILFLGGSQGAKAINEFALLNAP 198

Qy     220 ELCENLVKDDQVQVVVVCGKNTALKESLSALEAENG--DKLKVLGYVERIDELFRITDCM 277
           :| :  :|     :  :|| |:   | :     | |  ||:::  :   | |:    |
Db     199 KLTKQGIK-----ITHICGPNS--YEQVRFFYQELGLLDKIELFAFHNNITEVMHRADLC 251

Qy     278 ITKPGGITLTEATAIGVPVILYKPVP----GQEKENANFFEDRGAAIVVNRHE----EIL 329
           ::: |  :: |  | |:| | : | |       :  |   ||      || ::|    ::
Db     252 VSRAGASSVWELCANGLPTI-FIPYPFASNNHQYYNVLEFEKENLCYVVPQNELLPKKLF 310

Qy     330 ESVTSLLADED----TLHRMKKNIKDLHLANSSEVILEDIL 366
           | :  |   :|      |  :   ::     : :: |:| ||
Db     311 EVIRKLNQKDDQGNKNLTTISAKLQQKIAKDGAKTIIETIL 351
```

```
RESULT 14
MURG_MYCTU
ID   MURG_MYCTU     STANDARD;      PRT;   410 AA.
AC   O06224;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG OR RV2153C OR MTCY270.15.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sqares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z95388; CAB08640.1; -.
DR   TubercuList; Rv2153c; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   410 AA;  41860 MW;  96902AFE356FC30B CRC64;


  Query Match             6.5%;  Score 128.5;  DB 1;  Length 410;
  Best Local Similarity   19.9%;  Pred. No. 0.071;
  Matches   54;  Conservative   55;  Mismatches  124;  Indels   39;  Gaps   11;

Qy      98 ELVDEHQPDIIINTFPMIVVPEY-----------RRRTGRVIPTFNVMTDFCLHKIWVH 145
           :::|:  |:::    : :| |            |||   ||   |      :::  |
Db     121 DVLDDVDADVVVGFGGYVALPAYLAARGLPLPPRRRRRIPVVIHEANARAGLA-NRVGAH 179

Qy     146 ENVDKYYVATDYVKEKLLEIGTHPSNVKITGIPIRPQ---FEESMPVGPIYKKYNLSPNK 202
                | |   : : |      :: |:|:|     : ::       :    :
Db     180 --------TADRVLSAVPDSGLR--RAEVVGVPVRASIAALDRAVLRAEARAHFGFPDDA 229

Qy     203 KVLLIMAGAHGVLKNVKELCENLVKDDQVQVVVVCGKNTALKESLSAL--EAENGDKLKV 260
           :|||:  |: | : ::         |  :    ||  :    :::  |   |:
Db     230 RVLLVFGGSQGAV-SLNRAVSGAAAD--LAAAGVCVLHAHGPQNVLELRRRAQGDPPYVA 286

Qy     261 LGYVERIDELFRITDCMITKPGGITLTEATAIGVPVILYKPVP---GQEKENANFFEDRG 317
           : |::|::  :   | :| : | :|: | :|:|:| | | |:|   |::: ||    : |
Db     287 VPYLDRMELAYAAADLVICRAGAMTVAEVSAVGLPAI-YVPLPIGNGEQRLNALPVVNAG 345

Qy     318 AAIVVN----RHEEILESVTSLLADEDTLHRM 345
             :||       | :   |  || |   |  |
Db     346 GGMVVADAALTPELVARQVAGLLTDPARLAAM 377


RESULT  15
NIFE_ANASP
ID   NIFE_ANASP     STANDARD;      PRT;   480 AA.
AC   Q44144;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   NITROGENASE IRON-MOLYBDENUM COFACTOR BIOSYNTHESIS PROTEIN NIFE.
GN   NIFE.
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Buikema W.J., Scappino L.A., Haselkorn R.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN THE BIOSYNTHESIS OF
CC       THE PROSTHETIC GROUP OF NITROGENASE (FEMO COFACTOR).
CC   -!- PATHWAY: FE-MO COFACTOR BIOSYNTHESIS.
CC   -!- SIMILARITY: BELONGS TO THE NIFD/NIFK/NIFE/NIFN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U47055; AAA87947.1; -.
DR   InterPro; IPR000318; -.
DR   InterPro; IPR000510; -.
DR   Pfam; PF00148; oxidored_nitro; 1.
DR   PROSITE; PS00699; NITROGENASE_1_1; 1.
DR   PROSITE; PS00090; NITROGENASE_1_2; 1.
KW   Nitrogen fixation.
SQ   SEQUENCE   480 AA;  52804 MW;  F4F1D1BC968665DF CRC64;


  Query Match             6.3%;  Score 123.5;  DB 1;  Length 480;
  Best Local Similarity   21.9%;  Pred. No. 0.19;
  Matches   60;  Conservative   38;  Mismatches  101;  Indels   75;  Gaps    9;

Qy      73 YYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEYRRRTGRVIPTFN 132
           :||:| : |:   ||  |:|:                        |: : || ::|
Db     272 FYGIDDM-NRCLRNIAAKLGD----------------------PDLQARTEKLIAEET 306

Qy     133 VMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKITGIPIRPQFEESMPVGPI 192
              |  |          :  : |  ||   :        :::     |   ||
Db     307 AALDLALAPYRARLKGKRVVLYTGGVKSWSIISAAKDLGIEVVATSTRKSTEE------- 359

Qy     193 YKKYNLSPNKKVLLIMAGAHGVL---KNVKELCENLVKDDQVQVVVVCGKN--TALKESL 247
                  :|  :  : || |::    | ||| : ||||  |  :::  |:|  ||||  :
Db     360 --------DKAKIKRLLGADGIMLEKGNAKELLQ-LVKDTQADMLIAGGRNQYTALKARI 410

Qy     248 SALEAENGDKLKVLGYVERID---ELFRITDCMITKPGGITLTEATAIGVPVILYKPVPG 304
             |:          |||  |:   ||:                          || |:
Db     411 PFLDINQERHHPYAGYVGMIEMARELYE----------------------ALYSPIWE 446

Qy     305 QEKENANFFEDRG--AAIVVNRHEEILESVTSLL 336
           | :: || : || |  |    : |: || |:  |:
Db     447 QIRKPAPWDEDMGILAHEYTSNHDHILASIEELI 480


Search completed: June 29, 2001, 09:04:10
Job time: 302 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 10:22:58 ; Search time 162.85 Seconds
(without alignments)
4430.205 Million cell updates/sec

Title: US-09-668-788-1
Perfect score: 1149
Sequence: 1 ttgaataccaataaaagagt.....aagccaaagtgctatcgtaa 1149

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1149	100.0	1149	20	AAZ23386	B. subtilis glycos
	2	120.8	10.5	936	22	AAF58252	Oligonucleotide D1
	3	120.8	10.5	936	22	AAF58254	Oligonucleotide D1
	4	120.8	10.5	936	22	AAF58257	Oligonucleotide D1
	5	120.8	10.5	936	22	AAF58259	Oligonucleotide D2
	6	120.8	10.5	936	22	AAF58262	Oligonucleotide D2
	7	120.8	10.5	938	22	AAF58255	Oligonucleotide D1
c	8	120	10.4	936	22	AAF58252	Oligonucleotide D1
c	9	120	10.4	936	22	AAF58254	Oligonucleotide D1
c	10	120	10.4	936	22	AAF58257	Oligonucleotide D1
c	11	120	10.4	936	22	AAF58259	Oligonucleotide D2
c	12	120	10.4	936	22	AAF58262	Oligonucleotide D2
c	13	120	10.4	938	22	AAF58255	Oligonucleotide D1
c	14	119.6	10.4	12658	18	AAV74438	Staphylococcus aur
	15	54.2	4.7	244	22	AAF58238	Oligonucleotide D1
c	16	53	4.6	244	22	AAF58238	Oligonucleotide D1
	17	48	4.2	1266	19	AAV12734	Monogalactosyldiac
	18	41.6	3.6	1676	19	AAV43014	Streptococcus pneu
	19	41.4	3.6	695	20	AAX20376	Borrelia burgdorfe
c	20	41	3.6	35515	20	AAX20252	Borrelia burgdorfe
	21	40.4	3.5	5998	20	AAX13056	Enterococcus faeca
c	22	39.8	3.5	10607	14	AAQ37894	Beta-casein coding
c	23	39.8	3.5	10607	19	AAV25594	Human beta-casein
c	24	39.6	3.4	452	18	AAV74893	Staphylococcus aur
c	25	39	3.4	4041	21	AAA70170	Plasmodium falcipa
	26	38.8	3.4	1311	18	AAV74805	Staphylococcus aur
	27	38.4	3.3	65632	21	AAA81502	N. meningitidis pa
c	28	38.4	3.3	349980	21	AAF21544	Neisseria meningit
c	29	38.2	3.3	2600	20	AAX08512	MurE gene. Staphy
c	30	37.8	3.3	296	21	AAC94179	Cat flea head and
	31	37.4	3.3	861	20	AAX99566	Nucleic acid seque
	32	37	3.2	2849	20	AAX20293	Borrelia burgdorfe
	33	36.8	3.2	1764	21	AAC47190	Arabidopsis thalia
	34	36.8	3.2	2982	17	AAT12999	Nitrate reductase
	35	36.6	3.2	1410	20	AAX61755	B. burgdorferi ant
c	36	36.6	3.2	910715	20	AAX20248	Borrelia burgdorfe
c	37	36.2	3.2	1029	20	AAX13600	Enterococcus faeca
c	38	36	3.1	2128	20	AAX76373	P. carinii lysyl-t
	39	35.6	3.1	909	22	AAC89872	Human zacrp7 degen
	40	35.4	3.1	2740	22	AAF32681	Human cDNA encodin
	41	35.4	3.1	3019	22	AAF62010	Tomato LeSNF4 (suc
c	42	35.4	3.1	9062	18	AAV74396	Staphylococcus aur
c	43	35.2	3.1	111309	20	AAX20250	Borrelia burgdorfe
	44	35	3.0	282	21	AAA97854	Human EST clone DN
c	45	35	3.0	3771	21	AAA10741	C. albicans gene C

ALIGNMENTS

RESULT 1
AAZ23386
ID AAZ23386 standard; DNA; 1149 BP.
XX
AC AAZ23386;
XX
DT 13-DEC-1999 (first entry)
XX
DE B. subtilis glycosyl transferase catalytic domain DNA.
XX
KW Catalytic domain; glycosyl transferase; processive activity; detergent;
KW oligosaccharide glycolipid; 3-oligoglucosyl-1,2-diacylglycerol;
KW food emulsifier; polymer modifier; glycosyl transferase; ss.
XX
OS Bacillus subtilis.
XX
PN DE19819958-A1.
XX
PD 30-SEP-1999.
XX
PF 05-MAY-1998; 98DE-1019958.
XX
PR 25-MAR-1998; 98DE-1013017.
XX
PA (GVSE-) GVS GES ERWERB & VERW LANDWIRTSCHAFTLICH.
PA (BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.
XX
PI Wolter FP, Jorasch P, Heinz E, Zaehringer U;
XX
DR WPI; 1999-552364/47.
DR P-PSDB; AAY33444.
XX
PT New glycosyl transferase protein, useful for producing glycolipids -
XX

PS Example 1; Page 6; 8pp; German.
XX
CC This invention describes a novel protein (I) comprising identical or
CC different catalytic domains of glycosyl transferases and which has
CC processive activity (i.e. builds up oligosaccharide glycolipids by
CC successive addition of glucose units). (I) can be used to produce
CC 3-oligoglucosyl-1,2-diacylglycerols useful as food emulsifiers, polymer
CC modifiers or detergents. This sequence encodes a Bacillus subtilis
CC glycosyl transferase catalytic domain which is used in the method of
CC the invention.
XX
SQ Sequence 1149 BP; 380 A; 217 C; 244 G; 308 T; 0 other;

Query Match 100.0%; Score 1149; DB 20; Length 1149;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ttgaataccaataaaagagtattaattttgactgcaaattacggaaatggacatgtgcag 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ttgaataccaataaaagagtattaattttgactgcaaattacggaaatggacatgtgcag 60

Qy     61 gtagccaaaacactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 gtagccaaaacactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct 120

Qy    121 aatttgtaccaagagtcaaatccgattgtttcagaggtaactcaataccttttatttaaaa 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 aatttgtaccaagagtcaaatccgattgtttcagaggtaactcaataccttttatttaaaa 180

Qy    181 agcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 agcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctat 240

Qy    241 aataaacgtaaattcaatatttacttttaaaatgggtaataaaagattgggcgaacttgtc 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 aataaacgtaaattcaatatttacttttaaaatgggtaataaaagattgggcgaacttgtc 300

Qy    301 gatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatac 360

Qy    361 agacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 agacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcat 420

Qy    421 aaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaa 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 aaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaa 480

Qy    481 aaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcagg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 aaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcagg 540

Qy    541 ccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcacca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcacca 600

Qy    601 aacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagag 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 aacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagag 660

Qy    661 ctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720

Qy    721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780

Qy    781 ctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgattaccaag 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 ctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgattaccaag 840

Qy    841 cccggcggcattactttgacagaagccacagccattggagtgcctgtcattctgtacaaa 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 cccggcggcattactttgacagaagccacagccattggagtgcctgtcattctgtacaaa 900

Qy    901 cccgtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcggagctgccatc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 cccgtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcggagctgccatc 960

Qy    961 gttgtgaaccgtcatgaagagattctcgagtcagtcacttcccttcttgcagatgaagat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 gttgtgaaccgtcatgaagagattctcgagtcagtcacttcccttcttgcagatgaagat 1020

Qy   1021 accttgcatcgcatgaagaaaaacattaaggaccttcatttagcaaactcctctgaagtg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 accttgcatcgcatgaagaaaaacattaaggaccttcatttagcaaactcctctgaagtg 1080

Qy   1081 attttagaggatatcctgaaggaatcagaaatgatgaccgccaaacaaaaagccaaagtg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 attttagaggatatcctgaaggaatcagaaatgatgaccgccaaacaaaaagccaaagtg 1140

Qy   1141 ctatcgtaa 1149
          |||||||||
Db   1141 ctatcgtaa 1149
```

RESULT 2
AAF58252
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 10.5%; Score 120.8; DB 22; Length 936;

Best Local Similarity 1.0%; Pred. No. 9.2e-24;
Matches 8; Conservative 483; Mismatches 295; Indels 0; Gaps 0;

```
Qy      1 ttgaataccaataaaagagtattaattttgactgcaaattacggaaatggacatgtgcag 60
Db      4 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 63

Qy     61 gtagccaaaacactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct 120
Db     64 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwww 123

Qy    121 aatttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaa 180
Db    124 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 183

Qy    181 agcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctat 240
Db    184 wwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 243

Qy    241 aataaacgtaaattcaatatttactttaaaatgggtaataaaagattgggcgaacttgtc 300
Db    244 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwww 303

Qy    301 gatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatac 360
Db    304 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 363

Qy    361 agacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcat 420
Db    364 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwww 423

Qy    421 aaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaa 480
Db    424 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 483

Qy    481 aaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcagg 540
Db    484 wwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 543

Qy    541 ccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcacca 600
Db    544 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwww 603

Qy    601 aacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagag 660
Db    604 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 663

Qy    661 ctgtgcgaaaaccttgtcaaggatgacgaagtgcaagtagttgtcgtgtgcgggaaaaat 720
Db    664 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwww 723

Qy    721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
Db    724 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 783

Qy    781 ctgggc 786
Db    784 wwwwgc 789
```

RESULT 3
AAF58254
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 10.5%; Score 120.8; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 9.2e-24;
Matches 8; Conservative 483; Mismatches 295; Indels 0; Gaps 0;

```
Qy      1 ttgaataccaataaaagagtattaattttgactgcaaattacggaaatggacatgtgcag 60
Db      4 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 63

Qy     61 gtagccaaaacactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct 120
Db     64 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwww 123

Qy    121 aatttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaa 180
Db    124 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 183

Qy    181 agcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctat 240
Db    184 wwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 243

Qy    241 aataaacgtaaattcaatatttactttaaaatgggtaataaaagattgggcgaacttgtc 300
Db    244 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwww 303

Qy    301 gatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatac 360
Db    304 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 363

Qy    361 agacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcat 420
Db    364 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwww 423

Qy    421 aaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaa 480
Db    424 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 483

Qy    481 aaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcagg 540
Db    484 wwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 543
```

```
Qy    541 ccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcacca 600
Db    544 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwww 603

Qy    601 aacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagag 660
Db    604 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 663

Qy    661 ctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720
Db    664 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 723

Qy    721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
Db    724 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 783

Qy    781 ctgggc 786
Db    784 wwwwgc 789
```

RESULT 4
AAF58257
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 10.5%; Score 120.8; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 9.2e-24;
Matches 8; Conservative 483; Mismatches 295; Indels 0; Gaps 0;

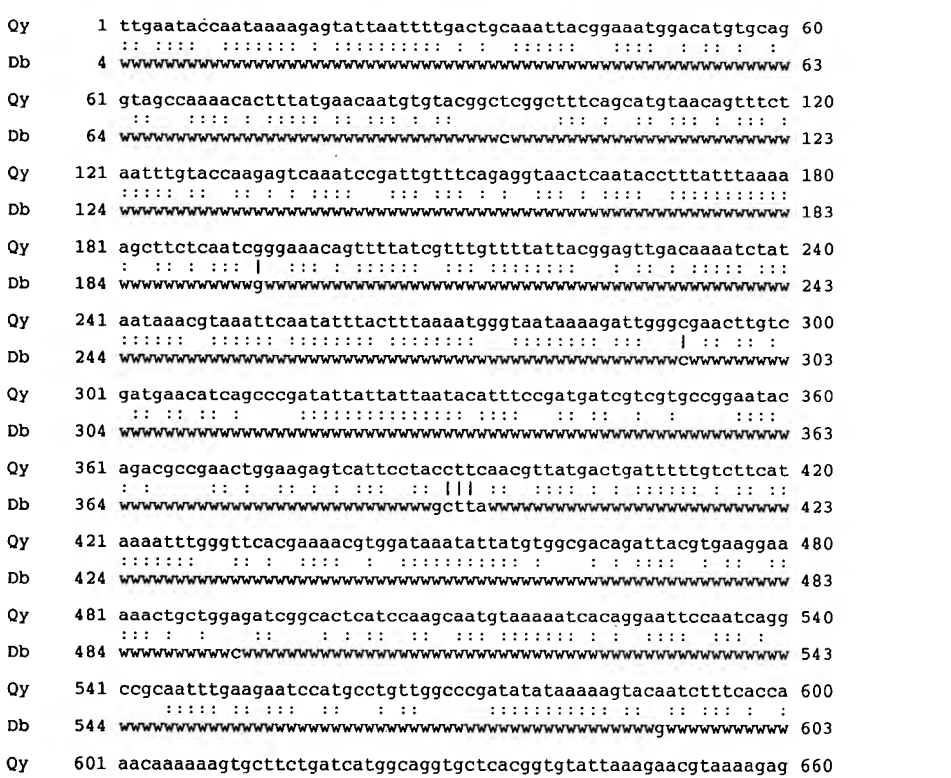
```
Qy      1 ttgaataccaataaaagagtattaattttgactgcaaattacggaaatggacatgtgcag 60
```

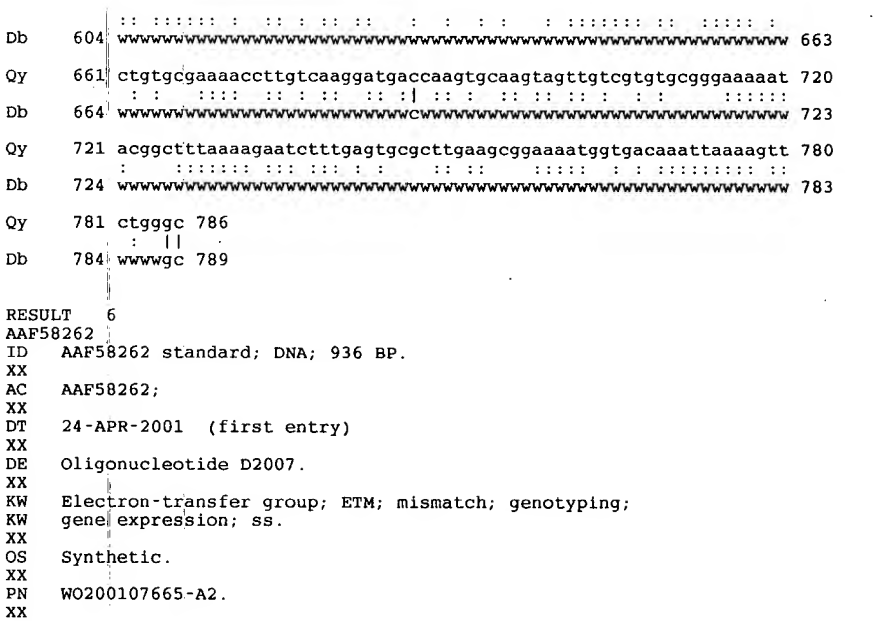
```
Db      4 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 63

Qy     61 gtagccaaaacactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct 120
Db     64 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwww 123

Qy    121 aatttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaa 180
Db    124 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 183

Qy    181 agcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctat 240
Db    184 wwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 243

Qy    241 aataaacgtaaattcaatatttactttaaaatgggtaataaaagattgggcgaacttgtc 300
Db    244 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwww 303

Qy    301 gatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatac 360
Db    304 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 363

Qy    361 agacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcat 420
Db    364 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwww 423

Qy    421 aaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaa 480
Db    424 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 483

Qy    481 aaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcagg 540
Db    484 wwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 543

Qy    541 ccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcacca 600
Db    544 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwww 603

Qy    601 aacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagag 660
Db    604 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 663

Qy    661 ctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720
Db    664 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 723

Qy    721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
Db    724 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 783

Qy    781 ctgggc 786
Db    784 wwwwgc 789
```

RESULT 5
AAF58259
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.

XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 10.5%; Score 120.8; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 9.2e-24;
Matches 8; Conservative 483; Mismatches 295; Indels 0; Gaps 0;





RESULT 6
AAF58262
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 10.5%; Score 120.8; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 9.2e-24;
Matches 8; Conservative 483; Mismatches 295; Indels 0; Gaps 0;

```
Qy    121 aatttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaa 180
Db    124 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 183

Qy    181 agcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctat 240
Db    184 wwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 243

Qy    241 aataaacgtaaattcaatatttacttttaaaatgggtaataaaagattgggcgaacttgtc 300
Db    244 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwww 303

Qy    301 gatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatac 360
Db    304 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 363

Qy    361 agacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcat 420
Db    364 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwww 423

Qy    421 aaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaa 480
Db    424 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 483

Qy    481 aaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcagg 540
Db    484 wwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 543

Qy    541 ccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcacca 600
Db    544 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwww 603

Qy    601 aacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagag 660
Db    604 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 663

Qy    661 ctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720
Db    664 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 723

Qy    721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
Db    724 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 783

Qy    781 ctgggc 786
Db    784 wwwwgc 789
```

RESULT 7
AAF58255
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1876.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

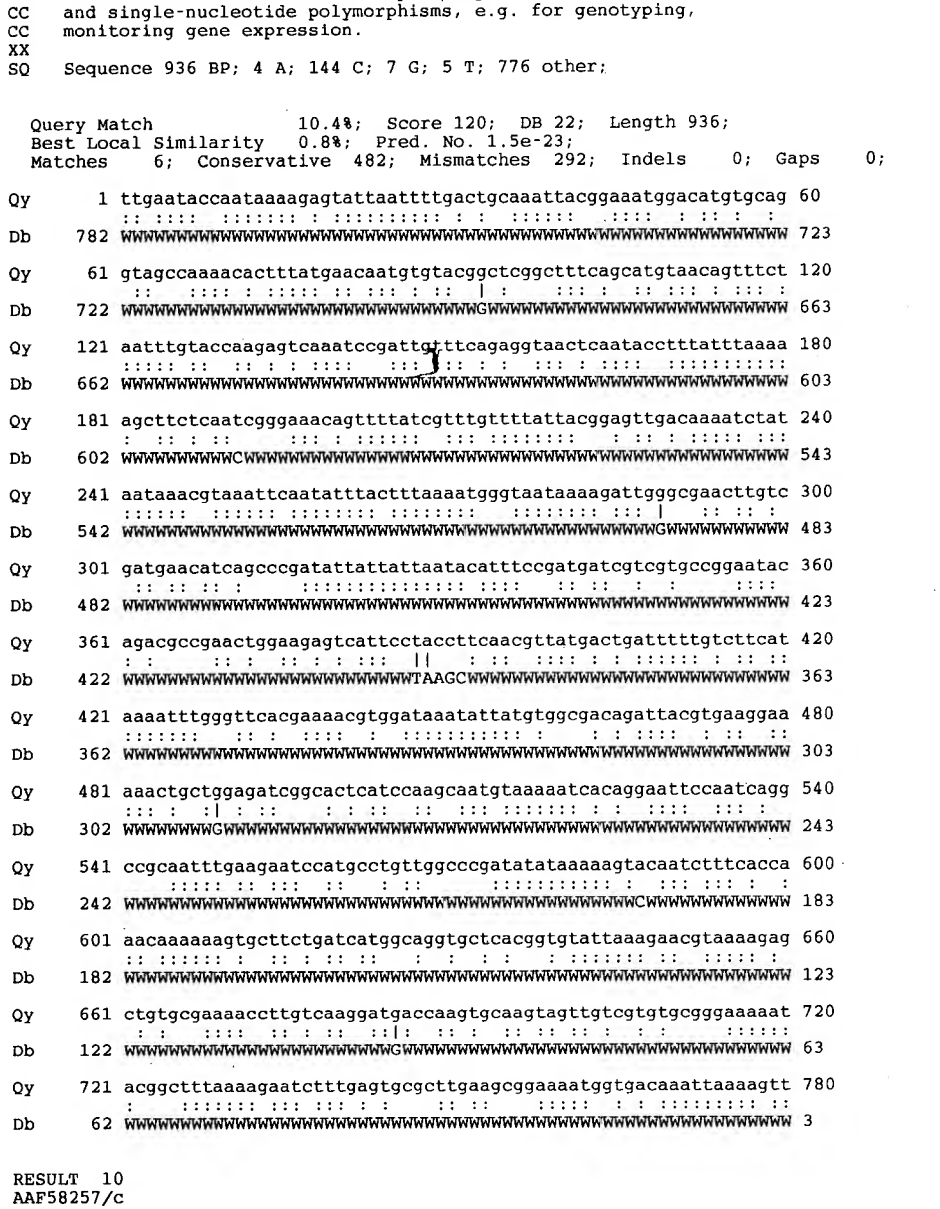
Query Match 10.5%; Score 120.8; DB 22; Length 938;
Best Local Similarity 1.0%; Pred. No. 9.2e-24;
Matches 8; Conservative 483; Mismatches 295; Indels 0; Gaps 0;

```
Qy      1 ttgaataccaataaaagagtattaattttgactgcaaattacggaaatggacatgtgcag 60
Db      4 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 63

Qy     61 gtagccaaaacactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct 120
Db     64 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwww 123

Qy    121 aatttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaa 180
Db    124 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 183

Qy    181 agcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctat 240
Db    184 wwwwwwwwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 243

Qy    241 aataaacgtaaattcaatatttacttttaaaatgggtaataaaagattgggcgaacttgtc 300
Db    244 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwww 303

Qy    301 gatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatac 360
Db    304 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 363

Qy    361 agacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcat 420
Db    364 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwww 423

Qy    421 aaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaa 480
Db    424 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 483

Qy    481 aaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcagg 540
Db    484 wwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 543

Qy    541 ccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcacca 600
Db    544 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwww 603

Qy    601 aacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagag 660
Db    604 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 663

Qy    661 ctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720
```

```
Db    664 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 723

Qy    721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
          :    ::::::: ::: ::: : :    :: ::    :::::  : : ::::::::: ::
Db    724 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 783

Qy    781 ctgggc 786
            :  ||
Db    784 wwwwgc 789


RESULT   8
AAF58252/c
ID   AAF58252 standard; DNA; 936 BP.
XX
AC   AAF58252;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1835.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;


  Query Match             10.4%;  Score 120;  DB 22;  Length 936;
  Best Local Similarity   0.8%;  Pred. No. 1.5e-23;
  Matches    6;  Conservative  482;  Mismatches  292;  Indels    0;  Gaps    0;

Qy      1 ttgaataccaataaaagagtattaattttgactgcaaattacggaaatggacatgtgcag 60
          :: ::::  ::::::: : :::::::::: : :  ::::::   ::::  : :: :  :
Db    782 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 723

Qy     61 gtagccaaaacactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct 120
          ::   :::: : ::::: :: ::: : ::  | :    ::: :  :: ::: : ::: :
Db    722 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwGwwwwwwwwwwwwwwwwwwwwwwwwwww 663

Qy    121 aatttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaa 180
          ::::: ::  :: : : ::::   ::: ::: : :  ::: : ::::  ::::::::::
Db    662 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 603

Qy    181 agcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctat 240
          :  :: : ::     ::: : ::::::  ::: ::::::::   : :: : ::::: :::
Db    602 wwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 543

Qy    241 aataaacgtaaattcaatatttactttaaaatgggtaataaaagattgggcgaacttgtc 300
          :::::::  :::::: :::::::: ::::::::   :::::::: ::: |   :: :: :
Db    542 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwGwwwwwwwwwww 483

Qy    301 gatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatac 360
          :: :: :: :     :::::::::::::::: ::::   :: ::  :  :      ::::
Db    482 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 423

Qy    361 agacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcat 420
          : :    :: :  :: : : :::  ||   : ::  :::: : : :::::: : :: ::
Db    422 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwTAAGCwwwwwwwwwwwwwwwwwwwwwwwwwww 363

Qy    421 aaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaa 480
          :::::::  :: :  ::::  :  :::::::::: :    : : ::::  : ::  ::
Db    362 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 303

Qy    481 aaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcagg 540
          ::: :  :| : ::    : : ::  ::  ::: :::::::: : :  :::: :: : :
Db    302 wwwwwwwwGwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 243

Qy    541 ccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcacca 600
              ::::: :: :::  ::   : ::       :::::::::: :  ::: ::: :  :
Db    242 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwCwwwwwwwwwwww 183

Qy    601 aacaaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagag 660
          :: :::::: :  :: : :: ::   :  : :  :  : :::::: ::  ::::: :
Db    182 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 123

Qy    661 ctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720
          : : :  ::::  :: : ::  ::|:  :: :  :: :: :: :  : :     ::::::
Db    122 wwwwwwwwwwwwwwwwwwwwwwwwwwGwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 63

Qy    721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
          :    ::::::: ::: ::: : :    :: ::    :::::  : : ::::::::: ::
Db     62 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 3


RESULT   9
AAF58254/c
ID   AAF58254 standard; DNA; 936 BP.
XX
AC   AAF58254;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1875.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
```

PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

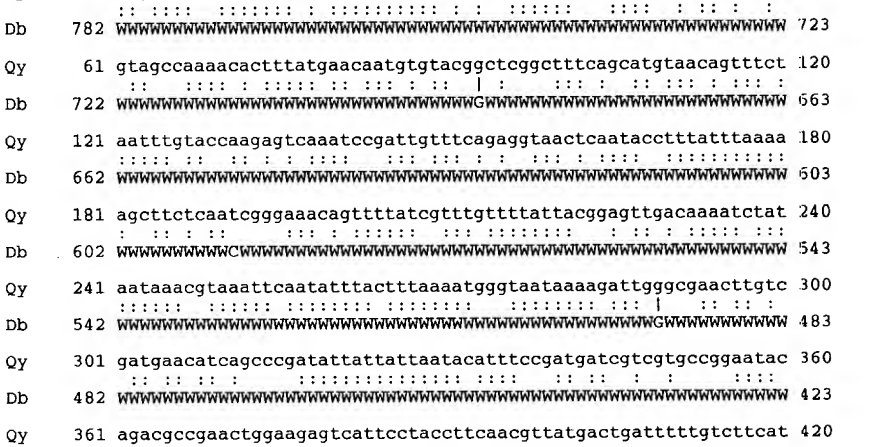
Query Match 10.4%; Score 120; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.5e-23;
Matches 6; Conservative 482; Mismatches 292; Indels 0; Gaps 0;

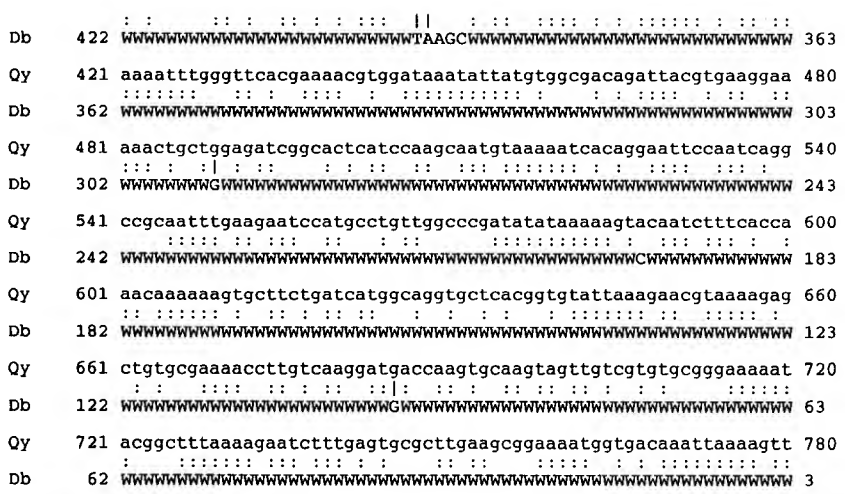


RESULT 10
AAF58257/c

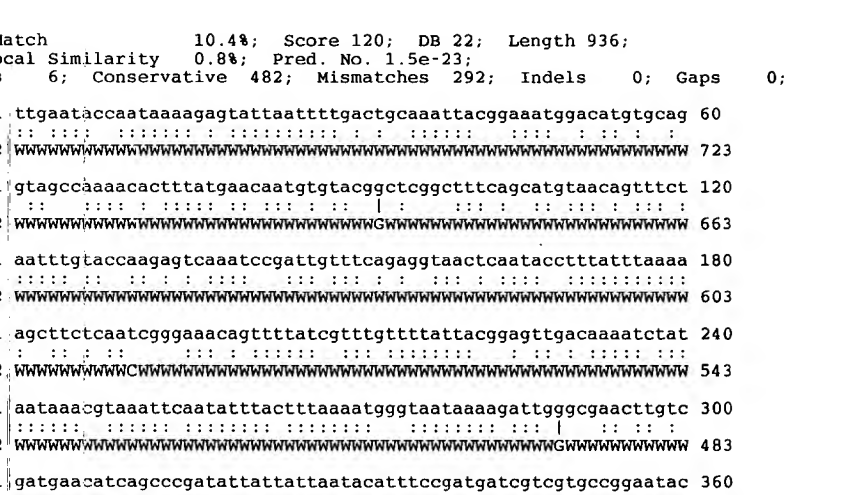
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 10.4%; Score 120; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.5e-23;
Matches 6; Conservative 482; Mismatches 292; Indels 0; Gaps 0;





RESULT 11
AAF58259/c
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;



RESULT 12
AAF58262/c
ID AAF58262 standard; DNA; 936 BP.
XX
AC AAF58262;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2007.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX

PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

Query Match 10.4%; Score 120; DB 22; Length 936;
Best Local Similarity 0.8%; Pred. No. 1.5e-23;
Matches 6; Conservative 482; Mismatches 292; Indels 0; Gaps 0;

Qy 1 ttgaataccaataaaagagtattaattttgactgcaaattacggaaatggacatgtgcag 60
Db 782 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 723

Qy 61 gtagccaaaacactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct 120
Db 722 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWW 663

Qy 121 aatttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaa 180
Db 662 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 603

Qy 181 agcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctat 240
Db 602 WWWWWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 543

Qy 241 aataaacgtaaattcaatatttactttaaaatgggtaataaaagattgggcgaacttgtc 300
Db 542 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWW 483

Qy 301 gatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatac 360
Db 482 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 423

Qy 361 agacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcat 420
Db 422 WWWWWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWWWWWWWWWWWWWWWWWWWWWWW 363

Qy 421 aaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaa 480
Db 362 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 303

Qy 481 aaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcagg 540
Db 302 WWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 243

Qy 541 ccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcacca 600
Db 242 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWWWWW 183

Qy 601 aacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagag 660
Db 182 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 123

Qy 661 ctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720
Db 122 WWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 63

Qy 721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
Db 62 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 3

RESULT 13
AAF58255/c
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1876.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 10.4%; Score 120; DB 22; Length 938;
Best Local Similarity 0.8%; Pred. No. 1.5e-23;
Matches 6; Conservative 482; Mismatches 292; Indels 0; Gaps 0;

Qy 1 ttgaataccaataaaagagtattaattttgactgcaaattacggaaatggacatgtgcag 60
Db 782 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 723

Qy 61 gtagccaaaacactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct 120
Db 722 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWW 663

```
Qy    121 aatttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaa 180
          ::::: ::  :: : : ::::   ::: ::: : :  ::: : ::::  ::::::::::
Db    662 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 603

Qy    181 agcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctat 240
          :  :: : ::     ::: : :::::: ::: ::::::::   : :: : ::::: :::
Db    602 WWWWWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 543

Qy    241 aataaacgtaaattcaatatttacttttaaaatgggtaataaaagattgggcgaacttgtc 300
          ::::::  :::::: :::::::: ::::::::   :::::::: ::: |   :: :: :
Db    542 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWW 483

Qy    301 gatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatac 360
           :: :: :: :     :::::::::::::::: ::::   :: ::  :  :     ::::
Db    482 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 423

Qy    361 agacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcat 420
          : :     :: :  :: : : :::  ||   : ::  :::: : : :::::: : :: ::
Db    422 WWWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWWWWWWWWWWWWWWWWWWWWWWWWW 363

Qy    421 aaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaa 480
          :::::::   :: :  ::::  :  :::::::::: :    : : ::::  : ::  ::
Db    362 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 303

Qy    481 aaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcagg 540
          ::: :  :| : ::    : : ::  ::  ::: ::::::: : :  ::::  ::: :
Db    302 WWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 243

Qy    541 ccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcacca 600
            ::::: :: :::  ::   : ::      :::::::::::: :  ::: ::: :  :
Db    242 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWWWW 183

Qy    601 aacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagag 660
          :: :::::: :  :: : :: ::   :  :  : :   : :::::: ::  ::::: :
Db    182 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 123

Qy    661 ctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720
           : :   ::::  :: : ::  ::|:  :: :  :: :: :: :  : :     ::::::
Db    122 WWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 63

Qy    721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780
          :    ::::::: ::: ::: : :    :: ::    :::::  : : :::::::::: ::
Db     62 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 3
```

RESULT 14
AAV74438/c
ID AAV74438 standard; DNA; 12658 BP.
XX
AC AAV74438;
XX
DT 16-MAR-1999 (first entry)
XX
DE Staphylococcus aureus contig SEQ ID #127.
XX
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
XX
OS Staphylococcus aureus.
XX
FH Key Location/Qualifiers
FT misc_feature 181..240
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 1981..2040
FT /*tag= b
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 3781..3840
FT /*tag= c
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 5581..5640
FT /*tag= d
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 7381..7440
FT /*tag= e
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 9181..9240
FT /*tag= f
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc_feature 10981..11040
FT /*tag= g
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
XX
PN EP786519-A2.
XX
PD 30-JUL-1997.
XX
PF 07-JAN-1997; 97EP-0100117.
XX
PR 05-JAN-1996; 96US-0009861.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
PI Rosen CA;
XX
DR WPI; 1997-374922/35.
XX
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
XX
PS Claim 1; Page 692-699; 3271pp; English.
XX
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
XX

```
SQ   Sequence 12658 BP; 4296 A; 1711 C; 2283 G; 3942 T; 426 other;


  Query Match             10.4%;  Score 119.6;  DB 18;  Length 12658;
  Best Local Similarity   49.5%;  Pred. No. 6.5e-23;
  Matches  367;  Conservative    0;  Mismatches  369;  Indels    6;  Gaps    2;

Qy     242 ataaacgtaaattcaatatttacttttaaaatgggtaataaaagattgggcgaacttgtcg 301
           ||||||   |      | |||||    |    ||   ||| |  ||     |  |  |
Db    5411 ATAAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTTATTGA 5352

Qy     302 atgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaataca 361
             || |  |||| ||| | ||  |    || ||||| |      ||  || ||| |   |
Db    5351 TAAAAGAAAAGCCAGATTTAATATTATTAACGTTTCCTACACCAGTTATGTCGGTACTAA 5292

Qy     362 gacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcata 421
            | |  |   |     |||||       |  || ||||| || | | |  | ||||
Db    5291 CTGAGCAATTTAACATTAATATTCCAGTTGCTACAGTGATGACAGACTATCGCTTACATA 5232

Qy     422 aaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaaa 481
           |||  ||| ||       |       | ||||||||||||| ||| |  |   |||  ||
Db    5231 AAAACTGGATTACGCCGTATTCAACAAGATATTATGTGGCAACAAAAGAAACGAAACAAG 5172

Qy     482 aactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcaggc 541
           |  |  | ||  | || | | ||||     |  || ||| | ||||| ||||| ||
Db    5171 ACTTCATAGACGTAGGTATTGATCCTTCAACAGTTAAAGTGACAGGTATTCCTATTGATA 5112

Qy     542 cgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcaccaa 601
           ||||||||    | ||   |     | |   |  | | |  ||||  |    |||
Db    5111 ACAAATTTGAAACGCCTATTAATCAAAAGCAGTGGTTAATAGACAACAACTTAGATCCAG 5052

Qy     602 acaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagagc 661
           | ||  ||    || | ||    || |||||    |||||||  ||     |  |   |
Db    5051 ATAAGCAAACTATTTTAATGTCAGCTGGTGCATTTGGTGTATCTAAAGGTTTTGACACGA 4992

Qy     662 tg---tgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaa 718
           ||      || |   | |  ||   ||   | |  |||||||||  |  | || || || |
Db    4991 TGATTACTGATATATTAGCGAAAAGTGCAAATGCACAAGTAGTTATGATTTGTGGTAAGA 4932

Qy     719 atacggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaag 778
             |  |   ||||    |||||  |  ||   | |      ||  |   || ||  ||
Db    4931 GCAAAGAGCTAAAGCGTTCTTTAACAGC---TAAGTTTAAATCTAACGAGAATGTATTGA 4875

Qy     779 ttctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgattacca 838
           ||||  || |||    | | |||  ||||     |       |   |    |||||||| |
Db    4874 TTCTAGGTTATACCAAACACATGAATGAATGGATGGCATCAAGTCAACTTATGATTACGA 4815

Qy     839 agcccggcggcattactttgacagaagccacagccattggagtgcctgtcattctgtaca 898
           | || || || || ||  | || ||||     |||  | |  | ||  | |||  |    |
Db    4814 AACCTGGTGGTATCACAATAACTGAAGGTTTCGCCCGTTGTATTCCAATGATTTTCCTAA 4755

Qy     899 aacccgtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcggagctgcca 958
           | || |  ||||| || ||    ||||||||    | |||||||||    ||   ||  |
Db    4754 ATCCTGCACCTGGTCAAGAGCTTGAAAATGCCCTTTACTTTGAAGAAAAAGGTTTTGGTA 4695

Qy     959 tcgttgtgaaccgtcatgaaga 980
             | |   |   ||  |||||
Db    4694 AAATCGCTGATACTCCAGAAGA 4673


RESULT  15
AAF58238
ID   AAF58238 standard; DNA; 244 BP.
XX
AC   AAF58238;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1250:D1102.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface -
XX
PS   Example 4; Page 120; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 244 BP; 19 A; 9 C; 12 G; 10 T; 194 other;


  Query Match              4.7%;  Score 54.2;  DB 22;  Length 244;
  Best Local Similarity    5.3%;  Pred. No. 1.4e-05;
  Matches   11;  Conservative  134;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     107 atgtaacagtttctaatttgtaccaagagtcaaatccgattgtttcagaggtaactcaat 166
           :: ::: : ::: :::::: ::  :: : : ::::   ::: ::: : :  ::: : :::
Db      13 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 72

Qy     167 acctttatttaaaaagcttctcaatcgggaaacagttttatcgtttgttttattacggag 226
           :  ::::::::::::  :: : :::    :::|: ::::::  ::: ::::::::   :
Db      73 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwww 132

Qy     227 ttgacaaaatctataataaacgtaaattcaatatttactttaaaatgggtaataaaagat 286
           :: : ::::: ::::::::::  :::::: :::::::: ::::::::   :::::::: ::
Db     133 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 192

Qy     287 tgggcgaacttgtcgatgaacatcagc 313
           :     :: :|||  |   ||| || |
Db     193 wwwwwwwwwwtgttaaaagacaccatc 219


Search completed: June 29, 2001, 11:49:30
Job time: 5192 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 09:32:17 ; Search time 1964.27 Seconds
(without alignments)
5529.448 Million cell updates/sec

Title: US-09-668-788-1
Perfect score: 1149
Sequence: 1 ttgaataccaataaaagagt.....aagccaaagtgctatcgtaa 1149

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	46.6	4.1	893	247	AZ679140	AZ679140 ENTJO94TR
c 2	45.4	4.0	740	219	CNS00AXB	AL056053 Drosophil
3	44.4	3.9	438	143	BE999232	BE999232 EST430955
c 4	44.2	3.8	548	120	AW787593	AW787593 945011D04
c 5	44.2	3.8	595	120	AW787592	AW787592 945011D04
c 6	44.2	3.8	610	104	AI979696	AI979696 614045D09
c 7	43.6	3.8	438	258	PT007G23U	AL447205 Parameciu
c 8	43.4	3.8	1101	219	CNS0039G	AL063921 Drosophil
c 9	43.4	3.8	1101	219	CNS0039Q	AL063931 Drosophil
c 10	43.2	3.8	1101	219	CNS016LI	AL106896 Drosophil
c 11	43	3.7	932	219	CNS017PY	AL108352 Drosophil
12	42.6	3.7	687	155	BG591315	BG591315 EST499157
13	42.4	3.7	852	247	AZ678871	AZ678871 ENTKE39TR
14	42	3.7	1001	219	CNS007BE	AL066953 Drosophil
15	41.6	3.6	526	21	AI487829	AI487829 EST246151
c 16	41.6	3.6	900	247	AZ667766	AZ667766 ENTLC14TR
c 17	41.4	3.6	1101	219	CNS00I9M	AL074705 Drosophil
18	41.2	3.6	569	223	AQ037612	AQ037612 CIT-HSP-2
19	41	3.6	519	250	AZ819893	AZ819893 2M0091I18
20	41	3.6	598	24	AI775222	AI775222 EST256322
21	41	3.6	692	230	AQ585760	AQ585760 RPCI-11-4
22	41	3.6	1001	219	CNS00GFA	AL072115 Drosophil
23	40.8	3.6	1101	219	CNS000D1	AL065414 Drosophil
c 24	40	3.5	843	219	CNS00CS1	AL059666 Drosophil
c 25	40	3.5	1101	219	CNS00Z3T	AL097187 Drosophil
26	39.6	3.4	234	220	CNS025K3	AL182172 Tetraodon
c 27	39.6	3.4	987	219	CNS014PQ	AL104456 Drosophil
28	39.4	3.4	495	105	AL368621	AL368621 MtBA25F05
29	39.2	3.4	711	251	AZ891758	AZ891758 RPCI-24-1
c 30	39.2	3.4	866	248	AZ684924	AZ684924 ENTMJ07TF
c 31	39.2	3.4	935	219	CNS00JKN	AL076398 Drosophil
c 32	39	3.4	419	107	AU034272	AU034272 AU034272
c 33	39	3.4	995	219	CNS005NG	AL060428 Drosophil
34	38.8	3.4	502	148	BF401494	BF401494 UI-R-CA0-
35	38.8	3.4	739	219	CNS015QP	AL105787 Drosophil
c 36	38.8	3.4	850	245	AZ550844	AZ550844 ENTGC80TF
c 37	38.8	3.4	909	220	CNS01ZJ5	AL174362 Tetraodon
38	38.8	3.4	928	219	CNS00DKY	AL071865 Drosophil
c 39	38.8	3.4	928	219	CNS00DKY	AL071865 Drosophil
c 40	38.8	3.4	1069	219	CNS0107G	AL098614 Drosophil
41	38.6	3.4	653	119	AW684087	AW684087 NF012B12N
42	38.6	3.4	858	248	AZ688017	AZ688017 ENTJW74TF
c 43	38.4	3.3	478	224	AQ124517	AQ124517 HS_3145_B
44	38.4	3.3	503	232	AQ677419	AQ677419 HS_5529_A
45	38.4	3.3	508	154	BG507491	BG507491 sac60a08.

ALIGNMENTS

```
RESULT  1
AZ679140/c
LOCUS       AZ679140      893 bp    DNA             GSS       14-DEC-2000
DEFINITION  ENTJO94TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ679140
VERSION     AZ679140.1  GI:11816286
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 893)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 15
            High quality sequence stop: 757.
FEATURES             Location/Qualifiers
     source          1. .893
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      326 a    120 c    137 g    310 t
ORIGIN
```

Query Match 4.1%; Score 46.6; DB 247; Length 893;
Best Local Similarity 50.7%; Pred. No. 0.091;
Matches 112; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```
Qy     117 ttctaatttgtaccaagagtcaaatccgattgtttcagaggtaactcaataccttttattt 176
           ||| ||| | ||  |  ||  |||||   ||||||  | |  ||    || | || || |
Db     467 TTCCAATATTTAGAACAAGCTAAATCATGTTGTTTAATATTTACTGGAAAATCTATACTA 408

Qy     177 aaaaagcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaat 236
             | |   ||  ||   |||||    | | |   ||  | ||  ||     |    |  |
Db     407 CCAGATTATCCTAAAATGGAAATTAGTATGTATATTCCTATACAACCTGACTCTGGATCT 348

Qy     237 ctataataaacgtaaattcaatatttactttaaaatgggtaataaaagattgggcgaact 296
           | || | ||     ||   ||  || ||| |  |||   |  | ||  ||  |   || |
Db     347 CCATTAAAACACATAAGCTAACTTTAACTCTTGAATATCTTTTCAACAATGTGTTTAATT 288

Qy     297 tgtcgatgaacatcagcccgatattattattaatacatttc 337
             |||  | ||| |  | |||  |  |||||||  ||||
Db     287 CAATGATTTATATCTGGTCAATAACAAAATTAATAATTTTC 247
```

```
RESULT  2
CNS00AXB/c
LOCUS       CNS00AXB      740 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR22J10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL056053
VERSION     AL056053.1  GI:4936923
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
```

```
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..740
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR22J10"
                     /note="end : T7"
BASE COUNT     206 a     16 c     25 g    188 t    305 others
ORIGIN

  Query Match             4.0%;  Score 45.4;  DB 219;  Length 740;
  Best Local Similarity   29.8%;  Pred. No. 0.18;
  Matches  150;  Conservative    8;  Mismatches  345;  Indels    0;  Gaps    0;

Qy    107 atgtaacagtttctaatttgtaccaagagtcaaatccgattgtttcagaggtaactcaat 166
Db    626 ATTWTANAATNNTTTTTNTTNATNANTAMAANANNTCTATTTTTNTTNNRNTNTNNCNNT 567

Qy    167 acctttatttaaaaagcttctcaatcgggaaacagttttatcgtttgttttattacggag 226
Db    566 WCAAAANNNNNNNANNNTGANAAAAAAAAAANNNAAATTTNTNTNTTTTTNTTTTNCMAAN 507

Qy    227 ttgacaaaatctataataaacgtaaattcaatatttactttaaaatgggtaataaaagat 286
Db    506 CTAAANGNANATTNATANNNNTTNTTTTTTNTNTNTTTTTTTTTTTTNHCAAAAAAAAAN 447

Qy    287 tgggcgaacttgtcgatgaacatcagcccgatattattattaatacatttccgatgatcg 346
Db    446 NANAAAANAAAATNNNNNNNNNTTATNNNNNTNNTTNTTTNNNNNNTTTTNANNTTNNN 387

Qy    347 tcgtgccggaatacagacgccgaactggaagagtcattcctaccttcaacgttatgactg 406
Db    386 TNNTACCAAACTAATNAANAANNARCAAANWANMAAATTNTTNTTTTNNNNTNNTTTTTN 327

Qy    407 atttttgtcttcataaaatttgggttcacgaaaacgtggataaatattatgtggcgacag 466
Db    326 NANNNTNTTTNTTNTNTNTNTNNNTNNNNNTNNNTNTANTNANTTTATATATTNGAAAAN 267

Qy    467 attacgtgaaggaaaaactgctggagatcggcactcatccaagcaatgtaaaaatcacag 526
Db    266 TANTANTNWATATAWTTTTAAAACNAAAAAAAATAATAAAAAAATAATTAAAAAAAANNAAG 207

Qy    527 gaattccaatcaggccgcaatttgaagaatccatgcctgttggcccgatatataaaaagt 586
Db    206 NNANNCCANAANNANNNNNNANANNNNNTCTBSTTNKGGBGGACCNAAMNAATTNNAANT 147

Qy    587 acaatctttcaccaaacaaaaaa 609
Db    146 NTTTTACWTTAAAAAATNANTNA 124

RESULT    3
BE999232
LOCUS       BE999232          438 bp    mRNA            EST       06-OCT-2000
DEFINITION  EST430955 GVSN Medicago truncatula cDNA clone pGVSN-15J11, mRNA
            sequence.
ACCESSION   BE999232
VERSION     BE999232.1  GI:10699508
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town
            ,C.D., Bowman,C.L., Craven,M.B., Cho,J. and Fraser,C.M.
  TITLE     ESTs from senescent nodules of Medicago truncatula
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Carroll P. Vance
            Department of Agronomy and Plant Genetics
            University of Minnesota
            411 Borlaug Hall, 1991 Upper Buford Circle, St.Paul, MN 55108 USA
            Tel: 612 625 5715
            Fax: 651-649-5058
            Email: vance004@maroon.tc.umn.edu
            University of Minnesota name: M27.3210e TIGR sequence name:
            MTKBA54TK More information is available at:
            http://chrysie.tamu.edu/medicago
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1..438
                     /organism="Medicago truncatula"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="pGVSN-15J11"
                     /clone_lib="GVSN"
                     /tissue_type="senescent root nodules"
                     /dev_stage="mixture of effective nodules from 40 day old
                     plants harvested 36 hours post shoot removal and nodules
                     collected from 2 month old plants at mid-pod stage"
                     /lab_host="E. coli strain SOLR"
                     /note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2:
                     XhoI; cDNA was prepared from polyA+ enriched RNA from the
                     mixture of effective nodules of 40 day old plants
                     harvested 36 hours post shoot removal and nodules
                     collected from 2 month old plants at mid-pod stage. The
                     cDNA was directionally ligated into the Uni-ZAP XR vector
                     from Stratagene and packaged using Gigapack III Gold
                     packaging extracts. Plasmids containing cDNA inserts were
                     excised from the recombinant lambda-ZAP phage using
                     Ex-Assist helper phage and propagated in SOLR cells."
BASE COUNT     151 a     76 c    114 g     97 t
ORIGIN

  Query Match             3.9%;  Score 44.4;  DB 143;  Length 438;
  Best Local Similarity   49.7%;  Pred. No. 0.29;
  Matches  145;  Conservative    0;  Mismatches  141;  Indels    6;  Gaps    1;

Qy    691 gtgcaagtagttgtcgtgtgcgggaaaaatacggctttaaaagaatctttgagtgcgctt 750
Db     87 GGGCAGATTGTAGTCATATGCGGTCGCAATAAGAACTTAGTAGCTTCTGTGGAAGCTATC 146

Qy    751 gaagcggaaaatggtgacaaattaaaagttctgggctatgtggagcgcattgatgagcta 810
Db    147 GAATGGAAAA------TACCAGTGAAGGTTAGAGGATTTGAGACAATTATGGCCAAATGG 200

Qy    811 tttcggatcacagattgcatgattaccaagcccggcggcattactttgacagaagccaca 870
Db    201 ATGGGAGCATGTGACTGCATCATATCAAAAGCTGGACCAGGAACAATTGCATAAGCATTG 260

Qy    871 gccattggagtgcctgtcattctgtacaaacccgtgcctggccaggaaaaagaaaatgca 930
Db    261 ATCACAGGTCTTCCTATAATCCTCAACGACTATATCCCTGGACAGGAAAAGGGTAATGTG 320

Qy    931 aacttctttgaagaccgcggagctgccatcgttgtgaaccgtcatgaagaga 982
```

```
Db    321 CCTTATGTGGTAAACAATGGAGCTGGTGTCTTCACTCGCAGTCCTAAAGAAA 372


RESULT   4
AW787593/c
LOCUS       AW787593      548 bp    mRNA            EST       16-MAY-2000
DEFINITION  945011D04.X2 945 - Mixed adult tissues from Walbot lab, same as 707
            (SK) Zea mays cDNA, mRNA sequence.
ACCESSION   AW787593
VERSION     AW787593.1  GI:7844371
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 945011  row: D  column: 04.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="945 Mixed adult tissues from Walbot lab,
                     same as 707 (SK)"
                     /tissue_type="tassel, kernal, silk, husk, root, leaf"
                     /dev_stage="fully-grown"
                     /lab_host="DH10B"
                     /note="Organ: tassel, kernal, silk, husk, root, leaf;
                     Vector: pGAD10; Site_1: EcoRI; cDNA library from fully
                     differentiated maize tissues from an active Mutator plant.
                     Tissue ratio is 4:2:1:1:1:1 (tassel, kernel, silk, husk,
                     root, leaf). Unidirectionally cloned.  New library number
                     given to library 707  for additional sequencing."
BASE COUNT      163 a    131 c     97 g    157 t
ORIGIN


  Query Match             3.8%;  Score 44.2;  DB 120;  Length 548;
  Best Local Similarity   49.8%;  Pred. No. 0.35;
  Matches  112;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    771 attaaaagttctgggctatgtggagcgcattgatgagctatttcggatcacagattgcat 830
          | |  | ||    |||| |||        || || ||     |  |       ||||| ||
Db    531 AGTTCAGGTGAAAGGCTTTGTTACAAAGATGGAAGAATGTATGGGTGCTTGTGATTGTAT 472

Qy    831 gattaccaagcccggcggcattactttgacagaagccacagccattggagtgcctgtcat 890
           ||||| ||| | ||      |||  |  |||| || |    |  |||  | ||  | ||
Db    471 CATTACAAAGGCAGGACCTGGTACAATTGCAGAGGCAATGATCCGTGGCTTACCAATTAT 412

Qy    891 tctgtacaaacccgtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcgg 950
          |||  |  |      |  |||| ||||||   |  ||||    || | |||  ||    ||
Db    411 TCTAAATGATTATATTGCTGGACAGGAAGCTGGCAATGTCCCCTACGTTGTTGAAAATGG 352

Qy    951 agctgccatcgttgtgaaccgtcatgaagagattctcgagtcagt 995
          |  ||  |   |  |||  ||  || |||||   ||  |||
Db    351 ATGTGGGAAGTTCTCGAAATCTCCAAAACAGATTGCGAAGATAGT 307


RESULT   5
AW787592/c
LOCUS       AW787592      595 bp    mRNA            EST       16-MAY-2000
DEFINITION  945011D04.X1 945 - Mixed adult tissues from Walbot lab, same as 707
            (SK) Zea mays cDNA, mRNA sequence.
ACCESSION   AW787592
VERSION     AW787592.1  GI:7844370
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 595)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 945011  row: D  column: 04.
FEATURES             Location/Qualifiers
     source          1. .595
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="945 - Mixed adult tissues from Walbot lab,
                     same as 707 (SK)"
                     /tissue_type="tassel, kernal, silk, husk, root, leaf"
                     /dev_stage="fully-grown"
                     /lab_host="DH10B"
                     /note="Organ: tassel, kernal, silk, husk, root, leaf;
                     Vector: pGAD10; Site_1: EcoRI; cDNA library from fully
                     differentiated maize tissues from an active Mutator plant.
                     Tissue ratio is 4:2:1:1:1:1 (tassel, kernel, silk, husk,
                     root, leaf). Unidirectionally cloned.  New library number
                     given to library 707  for additional sequencing."
BASE COUNT      176 a    142 c    104 g    173 t
ORIGIN


  Query Match             3.8%;  Score 44.2;  DB 120;  Length 595;
  Best Local Similarity   49.8%;  Pred. No. 0.36;
  Matches  112;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    771 attaaaagttctgggctatgtggagcgcattgatgagctatttcggatcacagattgcat 830
          | |  | ||    |||| |||        || || ||     |  |       ||||| ||
Db    528 AGTTCAGGTGAAAGGCTTTGTTACAAAGATGGAAGAATGTATGGGTGCTTGTGATTGTAT 469

Qy    831 gattaccaagcccggcggcattactttgacagaagccacagccattggagtgcctgtcat 890
           ||||| ||| | ||      |||  |  |||| || |    |  |||  | ||  | ||
Db    468 CATTACAAAGGCAGGACCTGGTACAATTGCAGAGGCAATGATCCGTGGCTTACCAATTAT 409

Qy    891 tctgtacaaacccgtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcgg 950
          |||  |  |      |  |||| ||||||   |  ||||    || | |||  ||    ||
Db    408 TCTAAATGATTATATTGCTGGACAGGAAGCTGGCAATGTCCCCTACGTTGTTGAAAATGG 349

Qy    951 agctgccatcgttgtgaaccgtcatgaagagattctcgagtcagt 995
          |  ||  |   |  |||  ||  || |||||   ||  |||
Db    348 ATGTGGGAAGTTCTCGAAATCTCCAAAACAGATTGCGAAGATAGT 304


RESULT   6
AI979696/c
LOCUS       AI979696      610 bp    mRNA            EST       30-AUG-1999
DEFINITION  614045D09.x4 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AI979696
VERSION     AI979696.1  GI:5804726
KEYWORDS    EST.
```

```
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614045  row: D  column: 09.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      182 a    145 c    107 g    175 t      1 others
ORIGIN
```

```
  Query Match             3.8%;  Score 44.2;  DB 104;  Length 610;
  Best Local Similarity   49.8%;  Pred. No. 0.36;
  Matches  112;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy    771 attaaaagttctgggctatgtggagcgcattgatgagctatttcggatcacagattgcat 830
          | |  | ||    |||| |||       || || ||     |  |       ||||| ||
Db    535 AGTTCAGGTGAAAGGCTTTGTTACAAAGATGGAAGAATGTATGGGTGCTTGTGATTGTAT 476

Qy    831 gattaccaagcccggcggcattactttgacagaagccacagccattggagtgcctgtcat 890
           ||||| ||| | ||      |||  |  |||| || |    |  |||  | ||  | ||
Db    475 CATTACAAAGGCAGGACCTGGTACAATTGCAGAGGCAATGATCCGTGGCTTACCAATTAT 416

Qy    891 tctgtacaaacccgtgcctggccaggaaaaagaaaatgcaaacttctttgaagaccgcgg 950
          |||  |  |     |  |||| ||||||   |  ||||    || | |||  ||    ||
Db    415 TCTAAATGATTATATTGCTGGACAGGAAGCTGGCAATGTCCCCTACGTTGTTGAAAATGG 356

Qy    951 agctgccatcgttgtgaaccgtcatgaagagattctcgagtcagt 995
          |  ||  |   |   |||   ||   || |||||    ||  |||
Db    355 ATGTGGGAAGTTCTCGAAATCTCCAAAACAGATTGCGAAGATAGT 311
```

```
RESULT   7
PT007G23U/c
LOCUS       PT007G23U     438 bp    DNA             GSS       02-NOV-2000
DEFINITION  Paramecium tetraurelia sequence M07D12u of the end of plasmid
            PT007G23, genomic survey sequence.
ACCESSION   AL447205
VERSION     AL447205.1  GI:11122266
KEYWORDS    GSS.
SOURCE      Paramecium tetraurelia.
  ORGANISM  Paramecium tetraurelia
            Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
            Paramecium.
REFERENCE   1  (bases 1 to 438)
  AUTHORS   Keller,A.M. and Cohen,J.
  TITLE     An indexed genomic library for Paramecium complementation cloning
  JOURNAL   J. Eukaryot. Microbiol. 47 (1), 1-6 (2000)
  MEDLINE   20114709
REFERENCE   2  (bases 1 to 438)
  AUTHORS   Dessen,P., Zagulski,M., Gromadka,R., Plattner,H., Kissmehl,R.,
            Meyer,E., Betermier,M., Schultz,J.E., Linder,J., Pearlman,R.E.,
            Kung,C., Forney,J., Satir,B., Van Houten,J.L., Keller,A.M.,
            Froissard,M., Sperling,L. and Cohen,J.
  TITLE     Paramecium genome survey: a pilot project
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 438)
  AUTHORS   Gromadka,R. and Zagulski,M.
  TITLE     Random sequencing of the Paramecium macronuclear genome
  JOURNAL   Unpublished
  REMARK    Institute of Biochemistry and Biophysics, Polish Academy of
            Sciences, Warsaw, Poland
REFERENCE   4  (bases 1 to 438)
  AUTHORS   Cohen,J. and Sperling,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2000) Paramecium Genome Survey Project, Centre de
            Genetique Moleculaire, Centre National de la Recherche
            Scientifique, 91198 Gif-sur-Yvette, France. E-mail:
            sperling@cgm.cnrs-gif.fr
COMMENT     The present survey of the Paramecium tetraurelia macronuclear
            genome consists of end sequences of a library of random 4-12 kb
            fragments obtained by Sau3A partial digestion of macronuclear DNA
            cloned in the BamHI site of pBSIIKS-. See [4].
            Genes are predicted from matches to other sequences. For more
            information about this sequence or the Paramecium Project, see
            http://caroll.vjf.inserm.fr/pt.
FEATURES             Location/Qualifiers
     source          1. .438
                     /organism="Paramecium tetraurelia"
                     /macronuclear
                     /strain="stock d4-2"
                     /db_xref="taxon:5888"
BASE COUNT      168 a     44 c     54 g    172 t
ORIGIN
```

```
  Query Match             3.8%;  Score 43.6;  DB 258;  Length 438;
  Best Local Similarity   52.8%;  Pred. No. 0.48;
  Matches   94;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy    105 gcatgtaacagtttctaatttgtaccaagagtcaaatccgattgtttcagaggtaactca 164
          | || ||  ||||||  |||| ||| || |      |  | ||  ||| | | |   |  |
Db    403 GAATTTATTAGTTTCCGATTTATACTAAAAAGTGTAAACAATCTTTTTAAATGATTCAAA 344

Qy    165 ataccttttatttaaaaagcttctcaatcgggaaacagttttatcgtttgttttattacgg 224
          |  |||  | |||  ||   |  || |  | ||| ||   |     || | ||| |
Db    343 AGTCCTAGAATTACCAATGATTGCATTTAGAAAATAGAACTTAAAATTCTCTTAATTTAT 284

Qy    225 agttgacaaaatctataataaacgtaaattcaatatttacttttaaaatgggtaataaa 282
          | ||   | | ||||  | || | ||| | |||| |   |||| ||     | ||||
Db    283 TTTAGATTTAGTATATATTTAAAGCAAAATTAATAATAGATTTACAAACTAAATTAAA 226
```

```
RESULT   8
CNS0039G/c
LOCUS       CNS0039G     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
```

```
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR08K10"
                     /note="end : TET3"
BASE COUNT      201 a     64 c    131 g    202 t    503 others
ORIGIN


  Query Match             3.8%;  Score 43.4;  DB 219;  Length 1101;
  Best Local Similarity   14.1%;  Pred. No. 0.7;
  Matches   67;  Conservative  213;  Mismatches  195;  Indels    0;  Gaps    0;

Qy      11 ataaaagagtattaattttgactgcaaattacggaaatggacatgtgcaggtagccaaaa 70
           :::::: : |::|:::::: : |    :| :   : |:: : | ::: ||: ::  | :|
Db    1043 WWWWWWWWATWDTWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDA 984

Qy      71 cactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttctaatttgtacc 130
             : |:   ::::::: :: : || ::: |: |:     ::::: : ::::: |
Db     983 DGAGRRDGGRKRKDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKD 924

Qy     131 aagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaaagcttctcaa 190
           :  :::  | :  :: :|::  :: ::|| : :  :::  :::|: |: :|   :  | ::
Db     923 RKADDDDGAGDKDDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWW 864

Qy     191 tcgggaaacagttttatcgtttgttttattacggagttgacaaaatctataataaaacgta 250
           : |  :|: : ::  |:  : :: :: : :  ::|  :|| |::  : ::|  :
Db     863 WWGWADADWWTWDAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARK 804

Qy     251 aattcaatatttactttaaaatgggtaataaaagattgggcgaacttgtcgatgaacatc 310
           :::  :|::: :| :: ::|||   :  |  :: :::: : : :  :   : :::: ::
Db     803 RRDRKRADDKRDAADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDD 744

Qy     311 agcccgatattattattaatacatttccgatgatcgtcgtgccggaatacagacgccgaa 370
           ::    | :::: |:: :::: :| :   : ::| :: : :  :: |:|  |      :::
Db     743 RDRAGTAGRKWRRTWKRRWKRRDTRWDDADADDTARDDRRRGDDGADAGKGKKTGRKRR 684

Qy     371 ctggaagagtcattcctaccttcaacgttatgactgatttttgtcttcataaaatttggg 430
             ::|  ::| |::   |  || :   ::::::  :::    :  || : |::::|::
Db     683 RRDRATWDRTDAWWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAWDWWTWKAW 624

Qy     431 ttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaaaaact 485
           ::   ::  : :  :::||:|:|:| : ::  | |:|: | :::|:  :|::: |
Db     623 DWAKWDWKTRADRWDRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWT 569


RESULT   9
CNS0039Q/c
LOCUS       CNS0039Q     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08M12 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063931
VERSION     AL063931.1  GI:4941788
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR08M12"
                     /note="end : TET3"
BASE COUNT      260 a    155 c     97 g    385 t    204 others
ORIGIN


  Query Match             3.8%;  Score 43.4;  DB 219;  Length 1101;
  Best Local Similarity   35.7%;  Pred. No. 0.7;
  Matches   99;  Conservative   43;  Mismatches  135;  Indels    0;  Gaps    0;

Qy     176 taaaaagcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaa 235
           :| |:|:      ||   | |:| |  :|||: |:|:|:||:|: : : |    | ||||
Db    1026 DAKARAKAAAAAAAAATTGAAWAAAAAWTTAKAGKTKGKTTKAKGRGRAAAAAAAAAAAA 967

Qy     236 tctataataaacgtaaattcaatatttactttaaaatgggtaataaaagattgggcgaac 295
             | || |:    :|: :   | |||  ||||||: :|::|| ||||  |||::
Db     966 AAAAAAAAAWTWWAWAWAWTTTTTTTTKTTTTAAARRRGKKAAAAAAAAWGTTGRKTTTTT 907

Qy     296 ttgtcgatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccgg 355
           || :  :| || |  :       :|  |   : | |  :||:   :: |:  :  |:  ||
Db     906 TTWWAARTAAAAAAAAWAWTTWTKAAAAAAGKAAGTTAWTTWAATWWTAWAWWAATKGDGG 847

Qy     356 aatacagacgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtc 415
             : :  |  :       :  || |   |:|  |      |   ||:: |   |||||
Db     846 GRARGGGGKTTTTTKWAAAAAAAAAAWTTTTTTTAAAATTTTWWWAAAAATTTTGWAW 787

Qy     416 ttcataaaatttgggttcacgaaaacgtggataaata 452
           ||   |||| ||   |:    ||||    | ||| |
Db     786 TTTTAAAAAATTTTTTWTTWAAAAAAAAAAAAAAAAA 750


RESULT  10
CNS016LI/c
LOCUS       CNS016LI     1101 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106896
VERSION     AL106896.1  GI:5624374
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
```

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/plasmid="pBeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN16D22"
/note="end : T7"
BASE COUNT 203 a 220 c 84 g 158 t 436 others
ORIGIN

Query Match 3.8%; Score 43.2; DB 219; Length 1101;
Best Local Similarity 31.2%; Pred. No. 0.79;
Matches 88; Conservative 72; Mismatches 121; Indels 1; Gaps 1;

```
Qy       2 tgaataccaataaaagagtattaattttgactgcaaattacggaaatggacatgtgcagg 61
Db    1060 TAMAAAWMAAAMWMAAWMAAWAWATTTAYATTTHWTAATHTAWTWYWCAATAHWWAWWTT 1001

Qy      62 tagccaaaacactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttcta 121
Db    1000 TATATWAATHTATTWATAHWTATYWTWTWHWATWYWTMTYWTATWAWAHAAAWATASTAA 941

Qy     122 atttgtaccaagagtcaaatccgattgtttcagaggtaactcaataccttttatttaaaaa 181
Db     940 AYWHATAATWAWAWATAATTWTWWTWTAWWTWTWAWTAATATWATWAAMTAATWWWWAWH 881

Qy     182 gcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctata 241
Db     880 AWWWTAYAWWWWWAWTATWWTATAWTAATAW-WATTWTAATTATAWTWAAAATATATAWT 822

Qy     242 ataaacgtaaattcaatatttactttaaaatgggtaataaaa 283
Db     821 WAAAAAAAAWATHAAWAAWAAATATWAAAAAAAAHMTAAAAW 780
```

RESULT 11
CNS017PY/c
LOCUS CNS017PY 932 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37G19 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL108352
VERSION AL108352.1 GI:5628656
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Plasmid Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 932)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .932
/organism="Drosophila melanogaster"
/plasmid="pBeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN37G19"
/note="end : SP6"
BASE COUNT 186 a 102 c 257 g 220 t 167 others
ORIGIN

Query Match 3.7%; Score 43; DB 219; Length 932;
Best Local Similarity 34.3%; Pred. No. 0.85;
Matches 83; Conservative 48; Mismatches 111; Indels 0; Gaps 0;

```
Qy      95 tcggctttcagcatgtaacagtttctaatttgtaccaagagtcaaatccgattgtttcag 154
Db     855 TAGKKTRKAARADGTTAATARKKTTWAAAWAAATDDATWTTTHTWWTTTGGGTRAWATTT 796

Qy     155 aggtaactcaatacctttatttaaaaagcttctcaatcgggaaacagttttatcgtttgt 214
Db     795 WTTTTNTWTTTTTTTTTTTTTTMTMTTTTTTTAMAAATGAAWAAKAATTTTTWHATWWAA 736

Qy     215 tttattacggagttgacaaaatctataataaacgtaaattcaatatttactttaaaatgg 274
Db     735 ARGTAARTTKTAAAAAAAAAWTATATAATATDAARAAATKWMATWWTTMTTTTWAAWTTT 676

Qy     275 gtaataaaagattgggcgaacttgtcgatgaacatcagcccgatattattattaatacat 334
Db     675 TAAATAWATTAWTKWTCTVGSCCMCCCCCCCCCCBCCHTCMTMTTWTAWWATMYTTTYTH 616

Qy     335 tt 336
Db     615 TT 614
```

RESULT 12
BG591315
LOCUS BG591315 687 bp mRNA EST 12-APR-2001
DEFINITION EST499157 P. infestans-challenged leaf Solanum tuberosum cDNA clone BPLI8C16 5' sequence, mRNA sequence.
ACCESSION BG591315
VERSION BG591315.1 GI:13609455
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 687)
AUTHORS Zhang,P., Hernandez,M., Tornqvist,C.-E, Wirtz,U., Loukoianov,A., Rangel,P., Haberlach,G.T., Cho,J., Chiemingo,A., Bougri,O., Buell,C.R., Ronning,C.M., Helgeson,J. and Baker,B.
TITLE Generation of ESTs from Potato Leaves Challenged with Phytophthora infestans, Incompatible Reaction
JOURNAL Unpublished (2000)
COMMENT Contact: Cathy Ronning
The Institute for Genomic Research
For clone info: please contact Research Genetics, Libraries Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: M13F-R.
FEATURES Location/Qualifiers
source 1. .687
/organism="Solanum tuberosum"

/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="BPLI8C16"
/clone_lib="P. infestans-challenged leaf"
/tissue_type="leaf"
/dev_stage="6 week old"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Whole plants were challenged with 450,000 sporangia/ml P. infestans US-1(US 940501) in Biotron (Madison, Wisconsin). Leaf tissue was collected at 1, 2, 5, 12, and 24 hours post-challenge and frozen in liquid nitrogen immediately upon removal. Kennebec plants showed no signs of HR. Katahdin plants (susceptible to P. infestans US-1) were used as controls and showed infection. NOTE: We cannot exclude the possibility that this sequence is actually derived from Phytophthora rather than potato."

BASE COUNT 201 a 129 c 178 g 179 t
ORIGIN

Query Match 3.7%; Score 42.6; DB 155; Length 687;
Best Local Similarity 54.9%; Pred. No. 1;
Matches 84; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
Qy    776 aagttctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgatta 835
          | ||   ||| | |||       || || ||     |  |       |||||||| ||||
Db    267 AGGTAAAGGGGTTTGTCACTAAAATGGAGGAATGCATGGGTGCTTGTGATTGCATTATTA 326

Qy    836 ccaagcccggcggcattactttgacagaagccacagccattggagtgcctgtcattctgt 895
          ||||| | ||       ||| |  | |||||||       |||| ||||| | |||||
Db    327 CCAAGGCTGGTCCAGGGACTATTGCGGAAGCCATGATTCGTGGACTGCCTATAATTCTTA 386

Qy    896 acaaacccgtgcctggccaggaaaaagaaaatg 928
          |     | |  |||| ||||||   | |||||
Db    387 ATGGTTTCATTGCTGGTCAGGAAGCTGGAAATG 419
```

RESULT 13
AZ678871
LOCUS AZ678871 852 bp DNA GSS 14-DEC-2000
DEFINITION ENTKE39TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION AZ678871
VERSION AZ678871.1 GI:11816017
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 852)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 22
High quality sequence stop: 788.
FEATURES Location/Qualifiers
source 1. .852
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."

BASE COUNT 352 a 113 c 108 g 279 t
ORIGIN

Query Match 3.7%; Score 42.4; DB 247; Length 852;
Best Local Similarity 51.6%; Pred. No. 1.2;
Matches 97; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

```
Qy    161 ctcaatacctttatttaaaaagcttctcaatcgggaaacagttttatcgtttgtttttatt 220
          || ||  | ||| |  ||| |   | | | || | | | | |||| |  |||  ||| ||
Db    581 CTTAAATCTTTTTTAAAAACACAATTTTATTCAGAAGATATTTTTTTGTTTTTGTTTTTT 640

Qy    221 acggagttgacaaaatctataataaacgtaaattcaatatttactttaaaatgggtaata 280
          |    |  | ||| | || || ||| | ||   |  |||||    ||||  |    ||
Db    641 AAATTTTAAAGAAATTTTAAAAGAAAATTTAAAGAAGAATTTATAAAAAAAAAGAACTTA 700

Qy    281 aaagattgggcgaacttgtcgatgaacatcagcccgatattattattaatacatttccga 340
          ||| |   ||      ||    ||    || |     ||  | ||  | ||  |||| ||
Db    701 AAAAAAAAGGAATTATTAAGAATATCAATGATTATCATTCTTTTCCTCATTTATTTTGGA 760

Qy    341 tgatcgtc 348
            ||| ||
Db    761 CTATCATC 768
```

RESULT 14
CNS007BE
LOCUS CNS007BE 1001 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR15H24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL066953
VERSION AL066953.1 GI:4945517
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1001)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1001
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR15H24"
/note="end : TET3"
BASE COUNT 159 a 148 c 151 g 212 t 331 others
ORIGIN

Query Match 3.7%; Score 42; DB 219; Length 1001;
Best Local Similarity 26.5%; Pred. No. 1.6;
Matches 71; Conservative 83; Mismatches 113; Indels 1; Gaps 1;

```
Qy    120 taatttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaa 179
          |||:| :: | |: | |:|||:   :|: :|||:: | :    |:::||  ::| :  |||
Db     42 TAAWTGSYTCAAWWAWTYAAAWTWAWTYMWTTYMAATKSYTTTYMWTAAWWWTTYMAAAA 101

Qy    180 aagcttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatcta 239
          |:  : |   :     ||:   |: :  | ::: ||:::  |  :: :| : |||:  :
Db    102 AWAWWACSKTWAWTAAAAWTTTTYCYCCCCYYYYTTWWWAAAWTSMMYTYMAAAAMSRWT 161

Qy    240 taataaacgtaaattcaatatttacttttaaaatgggtaa-taaaagattgggcgaacttg 298
          |:: :|:|  |:  : :| : |:| ||:||:     ::: |:||:  ||:|| | :  :
Db    162 TWRGRAMCMAAMCMWTWACMATWAWTTWAAMCMAAAWWMSTWAAMYTTTKGGGGYWAAWT 221

Qy    299 tcgatgaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaat 358
          :: :  ::: : : |:| :|     || ::| : ::     || | , :|: |::  :|
Db    222 WMMMCYMWYYWAMCCSCMWTTARGGTAAWWAAWWWWAAWWAATWAAAWYCSGGYYCCMAG 281

Qy    359 acagacgccgaactggaagagtcattcc 386
           :  : | : :   :|   | :  :|:|
Db    282 GMCMWKGGYWWKGGRGGGAAAWTCYTYC 309
```

RESULT 15
AI487829
LOCUS AI487829 526 bp mRNA EST 29-JUN-1999
DEFINITION EST246151 tomato ovary, TAMU Lycopersicon esculentum cDNA clone cLED13F18, mRNA sequence.
ACCESSION AI487829
VERSION AI487829.1 GI:4383200
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 526)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato carpel tissue
JOURNAL Unpublished (1999)
COMMENT Contact: David Frisch
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 4366
Fax: 864 656 4293
Email: dfrisch@CLEMSON.EDU.
FEATURES Location/Qualifiers
source 1. .526
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLED13F18"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XL1-Blue MRF'"
/note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2: Xho1; cLED - Tomato Carpel EST Library. OligodT-primed and directionally cloned cDNA in vector Lamda ZAP II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."
BASE COUNT 148 a 95 c 143 g 138 t 2 others
ORIGIN

Query Match 3.6%; Score 41.6; DB 21; Length 526;
Best Local Similarity 54.2%; Pred. No. 1.7;
Matches 83; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
Qy    776 aagttctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgatta 835
          | ||   ||| | |||       || || ||     |  |       |||||||| ||||
Db    327 AGGTAAAGGGGTTTGTCACTAAAATGGAGGAATGCATGGGTGCTTGTGATTGCATTATTA 386

Qy    836 ccaagcccggcggcattactttgacagaagccacagccattggagtgcctgtcattctgt 895
          | ||| | ||       ||| |  | |||||||       |||| ||||| | |||||
Db    387 CTAAGGCTGGTCCGGNGACTATTGCGGAAGCCATGATTCGTGGACTGCCTATAATTCTTA 446

Qy    896 acaaacccgtgcctggccaggaaaaagaaaatg 928
          |      | |  |||| ||||||   | |||||
Db    447 ATGGTTACATTGCTGGTCAGGAAGCTGGAAATG 479
```

Search completed: June 29, 2001, 10:55:47
Job time: 5010 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 09:56:28 ; Search time 119.12 Seconds
(without alignments)
1786.972 Million cell updates/sec

Title: US-09-668-788-1
Perfect score: 1149
Sequence: 1 ttgaataccaataaaagagt.....aagccaaagtgctatcgtaa 1149

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	49	4.3	7218	1	US-08-232-463-14	Sequence 14, Appl
c	2	39.8	3.5	10607	1	US-08-078-090-3	Sequence 3, Appli
c	3	38.2	3.3	2600	3	US-08-988-251-1	Sequence 1, Appli
c	4	38.2	3.3	2600	4	US-09-386-048-1	Sequence 1, Appli
c	5	36	3.1	2128	2	US-08-415-593-39	Sequence 39, Appl
	6	34.6	3.0	486	1	US-08-828-511-1	Sequence 1, Appli
c	7	34.4	3.0	1991	2	US-08-415-593-40	Sequence 40, Appl
	8	34.2	3.0	5852	1	US-07-867-106-2	Sequence 2, Appli
c	9	34	3.0	1000	1	US-08-599-252-96	Sequence 96, Appl
c	10	34	3.0	1000	5	PCT-US96-06352-96	Sequence 96, Appl
c	11	34	3.0	1000	5	PCT-US96-06583-96	Sequence 96, Appl
c	12	33.8	2.9	1395	1	US-07-991-867B-25	Sequence 25, Appl
c	13	33.8	2.9	1395	1	US-08-107-755A-25	Sequence 25, Appl
c	14	33.8	2.9	1395	2	US-08-544-332-25	Sequence 25, Appl
c	15	33.8	2.9	6768	1	US-08-107-755A-1	Sequence 1, Appli
c	16	33.8	2.9	8457	1	US-07-991-867B-1	Sequence 1, Appli
c	17	33.8	2.9	8457	2	US-08-544-332-1	Sequence 1, Appli
c	18	32.8	2.9	174	2	US-08-378-235B-8	Sequence 8, Appli
c	19	32.8	2.9	296	4	US-09-235-614-38	Sequence 38, Appl
	20	32.8	2.9	615	4	US-08-998-416-186	Sequence 186, App
	21	32.8	2.9	1107	1	US-08-165-038-3	Sequence 3, Appli
	22	32.8	2.9	1107	1	US-08-621-081A-3	Sequence 3, Appli
	23	32.8	2.9	1107	2	US-08-876-781-3	Sequence 3, Appli
	24	32.8	2.9	1384	4	US-09-058-489-30	Sequence 30, Appl
	25	32.8	2.9	2945	6	5196526-2	Patent No. 5196526
c	26	32.8	2.9	7560	4	US-09-103-478-4	Sequence 4, Appli
	27	32.4	2.8	3350	4	US-08-617-860B-3	Sequence 3, Appli
	28	32.4	2.8	9048	4	US-08-973-273-4	Sequence 4, Appli
c	29	32.2	2.8	389	3	US-08-851-843A-62	Sequence 62, Appl
c	30	32.2	2.8	389	4	US-08-974-549A-8	Sequence 8, Appli
c	31	32.2	2.8	2171	3	US-08-851-843A-100	Sequence 100, App
c	32	32.2	2.8	2171	4	US-08-974-549A-266	Sequence 266, App
c	33	32.2	2.8	2176	4	US-08-974-549A-3	Sequence 3, Appli
c	34	32.2	2.8	3855	4	US-08-974-549A-4	Sequence 4, Appli
c	35	32.2	2.8	4015	3	US-08-851-843A-224	Sequence 224, App
c	36	32.2	2.8	4015	4	US-08-974-549A-1	Sequence 1, Appli
c	37	32.2	2.8	4029	3	US-08-851-843A-173	Sequence 173, App
c	38	32.2	2.8	4029	4	US-08-974-549A-292	Sequence 292, App
c	39	32.2	2.8	4037	4	US-08-974-549A-343	Sequence 343, App
	40	32.2	2.8	6669	4	US-09-212-971-5	Sequence 5, Appli
	41	32.2	2.8	6669	4	US-08-800-929A-5	Sequence 5, Appli
	42	32	2.8	5496	1	US-08-181-629A-2	Sequence 2, Appli
c	43	31.8	2.8	2435	4	US-09-306-593-1	Sequence 1, Appli
c	44	31.6	2.8	7101	1	US-08-480-604A-9	Sequence 9, Appli
c	45	31.6	2.8	7101	2	US-08-405-496A-9	Sequence 9, Appli

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14
```

```
  Query Match             4.3%;  Score 49;  DB 1;  Length 7218;
  Best Local Similarity   2.7%;  Pred. No. 0.00047;
  Matches   10;  Conservative  216;  Mismatches  151;  Indels    0;  Gaps    0;

Qy     419 ataaaatttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaagg 478
           | || | || ||| |: ::::: : ::: ::: :  : : :: :: :::  : : :::::
Db    1449 AGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1390

Qy     479 aaaaactgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatca 538
           :::::  :  :::::  :: :  :   :::: :: : :::::  : :::::    ::  :
Db    1389 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1330

Qy     539 ggccgcaatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcac 598
           ::  : ::   ::::::   : :   :  ::   :: : : :::::: : ::      :
Db    1329 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1270

Qy     599 caaacaaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaag 658
            ::: :::::::: :     ::  : :: ::: :   : :: : :  :::::: : :::::
Db    1269 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1210

Qy     659 agctgtgcgaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaa 718
           ::  : : :::::    :  ::::: ::  ::: : ::: ::  :  : : : :::::::
Db    1209 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1150

Qy     719 atacggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaag 778
           : : ::    :::::::     ::: : :   :::: :::::: :: :: :::  :::::
Db    1149 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1090

Qy     779 ttctgggctatgtggag 795
             :::  : : ::::
Db    1089 RRRRRRRRRRRRRRRR 1073
```

```
RESULT   2
US-08-078-090-3/c
; Sequence 3, Application US/08078090
; Patent No. 5739407
;  GENERAL INFORMATION:
;    APPLICANT:  BERGSTROEM, SVEN
;    APPLICANT:  HERNELL, OLLE
;    APPLICANT:  LOENNERDAL, BO
;    APPLICANT:  HJALMARSSON, KARIN
;    APPLICANT:  HANSSON, LENNART
;    APPLICANT:  TOERNELL, JAN
;    APPLICANT:  STROEMQUIST, MATS
;    TITLE OF INVENTION:  HUMAN BETA-CASEIN PROCESS FOR PRODUCING
;    TITLE OF INVENTION:  IT AND USE THEREOF
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 SEVENTH STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/078,090
;      FILING DATE:  19930618
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/DK92/00236
;      FILING DATE:  19-AUG-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/DK91/00233
;      FILING DATE:  19-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  COOPER, IVER P.
;      REGISTRATION NUMBER:  28,005
;      REFERENCE/DOCKET NUMBER:  BERGSTROM2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)628-5197
;      TELEFAX:  (202)737-3528
;      TELEX:  248633
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10607 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(4804..4854, 5720..5746, 6726..6746, 6845
;      LOCATION:  ..6886, 7991..8521, 9440..9445)
US-08-078-090-3


  Query Match             3.5%;  Score 39.8;  DB 1;  Length 10607;
  Best Local Similarity   49.8%;  Pred. No. 0.19;
  Matches  101;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy     149 tttcagaggtaactcaataccttttatttaaaaagcttctcaatcgggaaacagttttatc 208
           ||| |||  |||   |     | |   |    || ||   ||       ||  |  | ||
Db    5489 TTTGAGAAATAAAAAATGTTATATGACTTTTGAGGTTTAAAAGGTAATTACTTTAATGTC 5430

Qy     209 gtttgtttttattacggagttgacaaaatctataataaacgtaaattcaatatttacttta 268
           |||  | ||          | |  | || | ||    || |  ||| || |   | |||
Db    5429 TTTTATATTTAGTAATCTATTAATATATTTTTATCTTACATTTATTAAAGACAGAGTTTT 5370

Qy     269 aaatgggtaataaaagattgggcgaacttgtcgatgaacatcagcccgatattattatta 328
             |||||||| | |||| |      | ||     | || ||       ||||||||||||
Db    5369 TTATGGGTAAAAGAAGAGTCACTTTAATTTGTATTAAATATTTTTATTATATTATTATTA 5310

Qy     329 atacatttccgatgatcgtcgtg 351
           ||| ||||  ||| ||  |  ||
Db    5309 ATATATTTATGATCATGATTATG 5287


RESULT   3
US-08-988-251-1/c
; Sequence 1, Application US/08988251
; Patent No. 6013497
;  GENERAL INFORMATION:
;    APPLICANT:  Wallis, Nicola G.
;    TITLE OF INVENTION:  MurE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  4000 Bell Atlantic Tower, 1717 Arch Stre
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  US
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/988,251
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dickinson, Todd Q
;      REGISTRATION NUMBER:  28,354
;      REFERENCE/DOCKET NUMBER:  GM10094
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-994-2252
```

```
;      TELEFAX:  215-994-2222
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2600 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-988-251-1


  Query Match             3.3%;  Score 38.2;  DB 3;  Length 2600;
  Best Local Similarity   46.9%;  Pred. No. 0.3;
  Matches  201;  Conservative    0;  Mismatches  213;  Indels   15;  Gaps    2;

Qy      10 aataaaagagtattaattttgactgcaaattacggaaatggacatgtgcaggtagccaaa 69
           |||||||   |||| ||| | ||||     | ||| || || ||| |||| ||  |  |
Db     545 AATAAAAAGATATTGATTATTACTGGCTCATTCGGTAACGGTCATATGCAAGTTACACAG 486

Qy      70 acactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct------aat 123
           |   |   | | |||  |       |     |  | ||| |||  ||   |       ||
Db     485 AGTATCGTTAATCAACTTAATGATATGAATCTAGACCATTTAAGCGTCATTGAGCACGAT 426

Qy     124 ttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaaagc 183
           || |     ||  |  |||| ||| | |     |  | || |    ||| | || |||
Db     425 TTATTTATGGAAGCTCATCCAATTTTGACTTCTATTTGTAAAAAATGGTATATCAATAGC 366

Qy     184 ttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctataat 243
           ||   |      | | |  | ||      |||||||          |||    |    ||
Db     365 TTTAAATATTTTAGAAATATGTACAAAGGGTTTTATT---------ACAGCCGCCCAGAT 315

Qy     244 aaacgtaaattcaatatttacttaaaatgggtaataaaagattgggcgaacttgtcgat 303
           ||||   |      | |||||    |    ||   ||| |  ||     |  |  |
Db     314 AAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTTATTGATA 255

Qy     304 gaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatacaga 363
           || |  |||| ||| | ||  |    || ||||| |     ||  || ||| |   |
Db     254 AAAGAAAAGCCAGATTTAATATTATTAACGTTTCCTACACCAGTTATGTCGGTACTAACT 195

Qy     364 cgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcataaa 423
              | |  |   |     |||||       |  || ||||| || | | |  | ||||||
Db     194 GAGCAATTTAACATTAATATTCCAGTTGCTACAGTGATGACAGACTATCGCTTACATAAA 135

Qy     424 atttgggtt 432
           |  |||  ||
Db     134 AACTGGATT 126


RESULT   4
US-09-386-048-1/c
; Sequence 1, Application US/09386048
; Patent No. 6103507
; GENERAL INFORMATION:
;  APPLICANT: Wallis, Nicola
;  APPLICANT: Shilling, Lisa
;  APPLICANT: Zabeierowski, Susan
;  TITLE OF INVENTION: MurE
;  FILE REFERENCE: GM10094-D1
;  CURRENT APPLICATION NUMBER: US/09/386,048
;  CURRENT FILING DATE: 1999-08-30
;  EARLIER APPLICATION NUMBER: US 08/988,251
;  EARLIER FILING DATE: 1997-12-10
;  EARLIER APPLICATION NUMBER: US 60/059,956
;  EARLIER FILING DATE: 1997-09-25
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2600
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-09-386-048-1


  Query Match             3.3%;  Score 38.2;  DB 4;  Length 2600;
  Best Local Similarity   46.9%;  Pred. No. 0.3;
  Matches  201;  Conservative    0;  Mismatches  213;  Indels   15;  Gaps    2;

Qy      10 aataaaagagtattaattttgactgcaaattacggaaatggacatgtgcaggtagccaaa 69
           |||||||   |||| ||| | ||||     | ||| || || ||| |||| ||  |  |
Db     545 AATAAAAAGATATTGATTATTACTGGCTCATTCGGTAACGGTCATATGCAAGTTACACAG 486

Qy      70 acactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct------aat 123
           |   |   | | |||  |       |     |  | ||| |||  ||   |       ||
Db     485 AGTATCGTTAATCAACTTAATGATATGAATCTAGACCATTTAAGCGTCATTGAGCACGAT 426

Qy     124 ttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaaagc 183
           || |     ||  |  |||| ||| | |     |  | || |    ||| | || |||
Db     425 TTATTTATGGAAGCTCATCCAATTTTGACTTCTATTTGTAAAAAATGGTATATCAATAGC 366

Qy     184 ttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctataat 243
           ||   |      | | |  | ||      |||||||          |||    |    ||
Db     365 TTTAAATATTTTAGAAATATGTACAAAGGGTTTTATT---------ACAGCCGCCCAGAT 315

Qy     244 aaacgtaaattcaatatttacttaaaatgggtaataaaagattgggcgaacttgtcgat 303
           ||||   |      | |||||    |    ||   ||| |  ||     |  |  |
Db     314 AAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTTATTGATA 255

Qy     304 gaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatacaga 363
           || |  |||| ||| | ||  |    || ||||| |     ||  || ||| |   |
Db     254 AAAGAAAAGCCAGATTTAATATTATTAACGTTTCCTACACCAGTTATGTCGGTACTAACT 195

Qy     364 cgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcataaa 423
              | |  |   |     |||||       |  || ||||| || | | |  | ||||||
Db     194 GAGCAATTTAACATTAATATTCCAGTTGCTACAGTGATGACAGACTATCGCTTACATAAA 135

Qy     424 atttgggtt 432
           |  |||  ||
Db     134 AACTGGATT 126


RESULT   5
US-08-415-593-39/c
; Sequence 39, Application US/08415593
; Patent No. 5912140
; Patent No. 5912140 5776726
;  GENERAL INFORMATION:
;    APPLICANT:  Whoriskey, Susan K.
;    APPLICANT:  Quinn, Cheryl L.
;    APPLICANT:  Tao, Niajun
;    APPLICANT:  Politis-Virk, Karen I.
;    APPLICANT:  Schimmel, Paul R.
;    TITLE OF INVENTION:  Recombinant Pneumoycstis Aminoacyl tRNA
;    TITLE OF INVENTION:  Synthetase Genes, Tester Strains and Assays
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/415,593
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E.
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  CPI94-09
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
```

```
;      TELEFAX:  617-862-9540
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2128 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-415-593-39


  Query Match             3.1%;  Score 36;  DB 2;  Length 2128;
  Best Local Similarity   49.5%;  Pred. No. 1.1;
  Matches   93;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy     146 ttgtttcagaggtaactcaatacctttatttaaaaagcttctcaatcgggaaacagtttt 205
           || |||    |  ||  |  |||  ||  |||| ||   ||   ||     ||| | |
Db    2107 TTTTTTTTTAAATACATACATAGATTATTTTACAAGAATTTATAACTTTAAAATAATCAC 2048

Qy     206 atcgtttgttttattacggagttgacaaaatctataataaacgtaaattcaatatttact 265
           |    | | || | |   | || |   | | ||||||| |  |   || ||| |||
Db    2047 TTTTAATATATTTTAAGTAATTTTAATCATTTTATAATACAAATTTGTTTAATTTTTTTA 1988

Qy     266 ttaaaatgggtaataaaagattgggcgaacttgtcgatgaacatcagcccgatattatta 325
           || |||    ||   ||| ||  |    |  | | ||  |  || |     | | ||||
Db    1987 TTCAAAAATATATGCAAAAATAAGAGTTATATTTTGAAAATAATTATTTTCAAAATATTT 1928

Qy     326 ttaataca 333
           ||| || |
Db    1927 TTATTAAA 1920


RESULT   6
US-08-828-511-1
; Sequence 1, Application US/08828511
; Patent No. 5731193
;  GENERAL INFORMATION:
;    APPLICANT:  MORI, Tetsuya
;    APPLICANT:  YAMAMOTO, Kozo
;    APPLICANT:  OHTA, Tsunetaka
;    TITLE OF INVENTION:  RECOMBINANT DNA AND TRANSFORMANT
;    TITLE OF INVENTION:  CONTAINING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BROWDY AND NEIMARK
;      STREET:  419 Seventh Street, N.W., Suite 300
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/828,511
;      FILING DATE:  31-MAR-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/355,245
;      FILING DATE:  09-DEC-1994
;      APPLICATION NUMBER:  JP 342237/1993
;      FILING DATE:  15-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  YUN, Allen C.
;      REGISTRATION NUMBER:  37,971
;      REFERENCE/DOCKET NUMBER:  MORI=39
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-828-511-1


  Query Match             3.0%;  Score 34.6;  DB 1;  Length 486;
  Best Local Similarity   51.6%;  Pred. No. 1.5;
  Matches   79;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     138 aaatccgattgtttcagaggtaactcaatacctttatttaaaaagcttctcaatcgggaa 197
           |||  | ||| |   | | ||  ||| |    | || |||  ||| |    ||   | |
Db      20 AAAATCTATTTTCAAAAAAGTTGCTCTAAGAATATAGTTATCAAGTTAAGTAAAATGTCA 79

Qy     198 acagttttatcgtttgttttattacggagttgacaaaatctataataaacgtaaattcaa 257
           | ||  || |  |||  ||| | |  |  ||  ||   | |  |||||    | ||| |
Db      80 ATAGCCTTTTAATTTAATTTTTAATTGTTTTATCATTCTTTGCAATAATAAAACATTAAC 139

Qy     258 tatttactttaaaatgggtaataaaagattggg 290
           | | ||||||  |||       || |  || | |
Db     140 TTTATACTTTTTAATTTAATGTATAGAATAGAG 172


RESULT   7
US-08-415-593-40/c
; Sequence 40, Application US/08415593
; Patent No. 5912140
; Patent No. 5912140 5776726
;  GENERAL INFORMATION:
;    APPLICANT:  Whoriskey, Susan K.
;    APPLICANT:  Quinn, Cheryl L.
;    APPLICANT:  Tao, Niajun
;    APPLICANT:  Politis-Virk, Karen I.
;    APPLICANT:  Schimmel, Paul R.
;    TITLE OF INVENTION:  Recombinant Pneumoycstis Aminoacyl tRNA
;    TITLE OF INVENTION:  Synthetase Genes, Tester Strains and Assays
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/415,593
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E.
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  CPI94-09
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-862-9540
;  INFORMATION FOR SEQ ID NO:  40:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1991 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  29..1780
US-08-415-593-40
```

```
  Query Match             3.0%;  Score 34.4;  DB 2;  Length 1991;
  Best Local Similarity   48.9%;  Pred. No. 3;
  Matches   92;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy     146 ttgtttcagaggtaactcaatacctttatttaaaaagcttctcaatcgggaaacagtttt 205
           || |||       ||  |  |||  ||  |||| ||   ||   ||     ||| | |
Db    1977 TTTTTTTTTTAATACATACATAGATTATTTTACAAGAATTTATAACTTTAAAATAATCAC 1918

Qy     206 atcgtttgttttattacggagttgacaaaatctataataaacgtaaattcaatatttact 265
           |    | | || | |   | || |   | | |||||||| |  |   || ||| |||
Db    1917 TTTTAATATATTTTAAGTAATTTTAATCATTTTATAATACAAATTTGTTTAATTTTTTTA 1858

Qy     266 ttaaaatgggtaataaaagattgggcgaacttgtcgatgaacatcagcccgatattatta 325
           || |||    ||   ||| ||  |    |  | | ||  |  || |     | | ||||
Db    1857 TTCAAAAATATATGCAAAAATAAGAGTTATATTTTGAAAATAATTATTTTCAAAATATTT 1798

Qy     326 ttaataca 333
           ||| || |
Db    1797 TTATTAAA 1790


RESULT   8
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
;  GENERAL INFORMATION:
;    APPLICANT:  Slade, Martin B
;    APPLICANT:  Chang, Andy C M
;    APPLICANT:  Williams, Keith L
;    TITLE OF INVENTION:  Improved Plasmid Vectors for Cellular
;    TITLE OF INVENTION:  Slime Moulds of the Genus Dictyostelium
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
;      STREET:  One Liberty Place 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/867,106
;      FILING DATE:  19920625
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  AU PJ 7187
;      APPLICATION NUMBER:  PCT/AU90/00530
;      FILING DATE:  02-NOV-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Feeney, Joanne Longo
;      REGISTRATION NUMBER:  35,134
;      REFERENCE/DOCKET NUMBER:  RICE-0002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5852 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2378..5038
US-07-867-106-2


  Query Match             3.0%;  Score 34.2;  DB 1;  Length 5852;
  Best Local Similarity   46.4%;  Pred. No. 5.3;
  Matches  111;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy     200 agttttatcgtttgttttattacggagttgacaaaatctataataaacgtaaattcaata 259
           | |||||   ||| |||| ||    || |  || |||  | ||| ||  |||| | || |
Db    1859 AATTTTAAAATTTTTTTTTTTTATTAGATCTCATAATTAAAAATCAATTTAAAATTAAAA 1918

Qy     260 tttacttttaaaatgggtaataaaagattgggcgaacttgtcgatgaacatcagcccgata 319
           ||| ||| ||||  | || ||           | |      | | || |       ||
Db    1919 GTTATTTTTAAATATGCAAAAACTATAAAAAACTAATGTAGTTTAACCAACTTTTTTCTA 1978

Qy     320 ttattattaatacatttccgatgatcgtcgtgccggaatacagacgccgaactggaagag 379
           ||  | ||  |   |||    | |  |  |    || | | |      ||    || |
Db    1979 TTTCTTTTTTTTTTTTTTTTTTTTTTTACTTTGAAAAAAAAAAAAAAAAAAAAAAAAAA 2038

Qy     380 tcattcctaccttcaacgttatgactgatttttgtcttcataaaatttgggttcacgaa 438
             |  ||  | ||  |  |  | | |  |||   | ||  | |  |||   || |  ||
Db    2039 AAAAACCCTCATTATAAATATTAATTACTTTGGTTTTTTTTGATTTTTTTTTTAATAAA 2097


RESULT   9
US-08-599-252-96/c
; Sequence 96, Application US/08599252
; Patent No. 5705343
;  GENERAL INFORMATION:
;    APPLICANT:  DRAYNA, DENNIS T.
;    APPLICANT:  FEDER, JOHN N.
;    APPLICANT:  GNIRKE, ANDREAS
;    APPLICANT:  KIMMEL, BRUCE E.
;    APPLICANT:  THOMAS, WINSTON J.
;    APPLICANT:  WOLFF, ROGER K.
;    TITLE OF INVENTION:  METHOD TO DIAGNOSE HEREDITARY
;    TITLE OF INVENTION:  HEMOCHROMATOSIS
;    NUMBER OF SEQUENCES:  124
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Ave. N.W., Suite 5500
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1888
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/599,252
;      FILING DATE:  09-FEB-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  9053-0001.21
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;      TELEX:  90-4030
;  INFORMATION FOR SEQ ID NO:  96:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1000 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-599-252-96


  Query Match             3.0%;  Score 34;  DB 1;  Length 1000;
```

```
  Best Local Similarity   49.4%;  Pred. No. 2.9;
  Matches   88;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy     157 gtaactcaatacctttatttaaaaagcttctcaatcgggaaacagttttatcgtttgttt 216
           || ||   | |  || ||| || ||  || |  |  | || | | | | |    || |
Db     242 GTGACATTAAATTTTAATTAAATAATATTTTATAATGTGACATATTATAAATTATTATAA 183

Qy     217 tattacggagttgacaaaatctataataaacgtaaattcaatatttactttaaaatgggt 276
           || ||   | |  |   |||  ||||||      | |  |   | |  || | ||    |
Db     182 TAATAAATAATATAAGTAATAAATAATATTTAATATTAAATATTATTTTTAATAAATAAT 123

Qy     277 aataaaagattgggcgaacttgtcgatgaacatcagcccgatattattattaatacat 334
           ||  ||| | ||    |   || |  | || || |    |   | |||| |||| |||
Db     122 AAATAAATAATGAATAATTATGACATTAAAAATAATAGAGTACTCATTACTAATCCAT 65


RESULT  10
PCT-US96-06352-96/c
; Sequence 96, Application PC/TUS9606352
;  GENERAL INFORMATION:
;    APPLICANT:  DRAYNA, DENNIS T.
;    APPLICANT:  FEDER, JOHN N.
;    APPLICANT:  GNIRKE, ANDREAS
;    APPLICANT:  KIMMEL, BRUCE E.
;    APPLICANT:  THOMAS, WINSTON J.
;    APPLICANT:  WOLFF, ROGER K.
;    TITLE OF INVENTION:  METHOD TO DIAGNOSE HEREDITARY
;    TITLE OF INVENTION:  HEMOCHROMATOSIS
;    NUMBER OF SEQUENCES:  124
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Ave. N.W., Suite 5500
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1888
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/06352
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/599,252
;      FILING DATE:  09-FEB-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  9053-0001.21
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;      TELEX:  90-4030
;  INFORMATION FOR SEQ ID NO:  96:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1000 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US96-06352-96


  Query Match             3.0%;  Score 34;  DB 5;  Length 1000;
  Best Local Similarity   49.4%;  Pred. No. 2.9;
  Matches   88;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy     157 gtaactcaatacctttatttaaaaagcttctcaatcgggaaacagttttatcgtttgttt 216
           || ||  . | |  || ||| || ||  || |  |  | || | | | | |    || |
Db     242 GTGACATTAAATTTTAATTAAATAATATTTTATAATGTGACATATTATAAATTATTATAA 183
```

```
Qy     217 tattacggagttgacaaaatctataataaacgtaaattcaatatttactttaaaatgggt 276
           || ||   | |  |   |||  ||||||      | |  |   | |  || | ||    |
Db     182 TAATAAATAATATAAGTAATAAATAATATTTAATATTAAATATTATTTTTAATAAATAAT 123

Qy     277 aataaaagattgggcgaacttgtcgatgaacatcagcccgatattattattaatacat 334
           ||  ||| | ||    |   || |  | || || |    |   | |||| |||| |||
Db     122 AAATAAATAATGAATAATTATGACATTAAAAATAATAGAGTACTCATTACTAATCCAT 65


RESULT  11
PCT-US96-06583-96/c
; Sequence 96, Application PC/TUS9606583
;  GENERAL INFORMATION:
;    APPLICANT:  DRAYNA, DENNIS T.
;    APPLICANT:  FEDER, JOHN N.
;    APPLICANT:  GNIRKE, ANDREAS
;    APPLICANT:  KIMMEL, BRUCE E.
;    APPLICANT:  THOMAS, WINSTON J.
;    APPLICANT:  WOLFF, ROGER K.
;    TITLE OF INVENTION:  METHOD TO DIAGNOSE HEREDITARY
;    TITLE OF INVENTION:  HEMOCHROMATOSIS
;    NUMBER OF SEQUENCES:  124
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Ave. N.W., Suite 5500
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1888
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/06583
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/599,252
;      FILING DATE:  09-FEB-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  9053-0001.21
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;      TELEX:  90-4030
;  INFORMATION FOR SEQ ID NO:  96:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1000 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US96-06583-96


  Query Match             3.0%;  Score 34;  DB 5;  Length 1000;
  Best Local Similarity   49.4%;  Pred. No. 2.9;
  Matches   88;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy     157 gtaactcaatacctttatttaaaaagcttctcaatcgggaaacagttttatcgtttgttt 216
           || ||   | |  || ||| || ||  || |  |  | || | | | | |    || |
Db     242 GTGACATTAAATTTTAATTAAATAATATTTTATAATGTGACATATTATAAATTATTATAA 183

Qy     217 tattacggagttgacaaaatctataataaacgtaaattcaatatttactttaaaatgggt 276
           || ||   | |  |   |||  ||||||      | |  |   | |  || | ||    |
Db     182 TAATAAATAATATAAGTAATAAATAATATTTAATATTAAATATTATTTTTAATAAATAAT 123

Qy     277 aataaaagattgggcgaacttgtcgatgaacatcagcccgatattattattaatacat 334
           ||  ||| | ||    |   || |  | || || |    |   | |||| |||| |||
Db     122 AAATAAATAATGAATAATTATGACATTAAAAATAATAGAGTACTCATTACTAATCCAT 65
```

```
RESULT 12
US-07-991-867B-25/c
; Sequence 25, Application US/07991867B
; Patent No. 5476781
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5476781el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/991,867B
;      FILING DATE:  12-DEC-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 92/14818
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,685
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1395 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-07-991-867B-25


  Query Match             2.9%;  Score 33.8;  DB 1;  Length 1395;
  Best Local Similarity   52.5%;  Pred. No. 3.8;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     204 ttatcgtttgttttattacggagttgacaaaatctataataaacgtaaattcaatattta 263
           | ||| |||   || || || | | || |||  || | |||||  ||||     |   ||
Db     628 TGATCATTTTACTTGTTGCGTAATAGATAAACACTTTGATAAAGATAAAAAAGCTGCATA 569

Qy     264 ctttaaaatgggtaataaaagattgggcgaacttgtcgatgaacatcagcccgatattat 323
           ||||   |   ||| |       |    |||||| |  |  || || |     ||||  |
Db     568 CTTTTTTAATAGTAGTGGTTATATACCAGAACTTATAAAACAAAATAAAAAATATATGTT 509

Qy     324 tattaatacatttccgatgat 344
           |||| |  |   |  ||  ||
Db     508 TATTGAATCTGATATGACTAT 488


RESULT 13
US-08-107-755A-25/c
; Sequence 25, Application US/08107755A
; Patent No. 5721352
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5721352el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  U.S.A.
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/107,755A
;      FILING DATE:  19-AUG-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,658
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1395 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-107-755A-25


  Query Match             2.9%;  Score 33.8;  DB 1;  Length 1395;
  Best Local Similarity   52.5%;  Pred. No. 3.8;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     204 ttatcgtttgttttattacggagttgacaaaatctataataaacgtaaattcaatattta 263
           | ||| |||   || || || | | || |||  || | |||||  ||||     |   ||
Db     628 TGATCATTTTACTTGTTGCGTAATAGATAAACACTTTGATAAAGATAAAAAAGCTGCATA 569

Qy     264 ctttaaaatgggtaataaaagattgggcgaacttgtcgatgaacatcagcccgatattat 323
           ||||   |   ||| |       |    |||||| |  |  || || |     ||||  |
Db     568 CTTTTTTAATAGTAGTGGTTATATACCAGAACTTATAAAACAAAATAAAAAATATATGTT 509

Qy     324 tattaatacatttccgatgat 344
           |||| |  |   |  ||  ||
Db     508 TATTGAATCTGATATGACTAT 488


RESULT 14
US-08-544-332-25/c
; Sequence 25, Application US/08544332
; Patent No. 5935777
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5935777el Entomopoxvirus Expression System
```

```
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gerard H. Bencen
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/544,332
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/991,867
;      FILING DATE:  07-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/107,755
;      FILING DATE:  19-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 92/14818
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,685
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bencen, Gerard H.
;      REGISTRATION NUMBER:  35,746
;      REFERENCE/DOCKET NUMBER:  UF114.C4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1395 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-544-332-25
```

```
  Query Match             2.9%;  Score 33.8;  DB 2;  Length 1395;
  Best Local Similarity   52.5%;  Pred. No. 3.8;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     204 ttatcgtttgttttattacggagttgacaaaatctataataaacgtaaattcaatattta 263
           | ||| |||   || || || | | || |||  || | |||||  ||||     |   ||
Db     628 TGATCATTTTACTTGTTGCGTAATAGATAAACACTTTGATAAAGATAAAAAAGCTGCATA 569

Qy     264 ctttaaaatgggtaataaaagattgggcgaacttgtcgatgaacatcagcccgatattat 323
           ||||   |   |||  |       |    |||||| |  |  || || |     ||||  |
Db     568 CTTTTTTAATAGTAGTGGTTATATACCAGAACTTATAAAACAAAATAAAAAATATATGTT 509

Qy     324 tattaatacatttccgatgat 344
           |||| |  |   |  ||  ||
Db     508 TATTGAATCTGATATGACTAT 488


RESULT  15
US-08-107-755A-1/c
; Sequence 1, Application US/08107755A
; Patent No. 5721352
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5721352el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  U.S.A.
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/107,755A
;      FILING DATE:  19-AUG-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,658
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6768 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Amsacta moorei entomopoxvirus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (65..1459)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1474..2151
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (2239..2475)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2502..2987
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3080..6091
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (6277..6768)
US-08-107-755A-1
```

```
  Query Match             2.9%;  Score 33.8;  DB 1;  Length 6768;
  Best Local Similarity   52.5%;  Pred. No. 7.2;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     204 ttatcgtttgttttattacggagttgacaaaatctataataaacgtaaattcaatattta 263
           | ||| |||   || || || | | || |||  || | |||||  ||||     |   ||
Db     692 TGATCATTTTACTTGTTGCGTAATAGATAAACACTTTGATAAAGATAAAAAAGCTGCATA 633

Qy     264 ctttaaaatgggtaataaaagattgggcgaacttgtcgatgaacatcagcccgatattat 323
           ||||   |   |||  |       |    |||||| |  |  || || |     ||||  |
Db     632 CTTTTTTAATAGTAGTGGTTATATACCAGAACTTATAAAACAAAATAAAAAATATATGTT 573
```

```
Qy     324 tattaatacatttccgatgat 344
           |||| |  |   |  ||  ||
Db     572 TATTGAATCTGATATGACTAT 552
```

Search completed: June 29, 2001, 11:46:43
Job time: 6615 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 29, 2001, 08:56:07 ; Search time 35.83 Seconds
(without alignments)
646.340 Million cell updates/sec

Title: US-09-668-788-2
Perfect score: 1970
Sequence: 1 MNTNKRVLILTANYGNGHVQ.....EDILKESEMMTAKQKAKVLS 382

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1970	100.0	382	20	AAY33444	B. subtilis glycos
2	326	16.5	522	21	AAB19039	Amino acid sequenc
3	309.5	15.7	492	21	AAG42414	Arabidopsis thalia
4	309.5	15.7	551	21	AAG42413	Arabidopsis thalia
5	305.5	15.5	422	19	AAW41148	Monogalactosyldiac
6	303.5	15.4	525	21	AAB19040	Amino acid sequenc
7	301.5	15.3	533	21	AAB19041	Amino acid sequenc
8	293	14.9	468	21	AAB19042	Amino acid sequenc
9	287.5	14.6	404	21	AAG42415	Arabidopsis thalia
10	114.5	5.8	468	20	AAW92952	WO9905287 Seq ID 1
11	113.5	5.8	355	21	AAY74401	Neisseria gonorrhe
12	112.5	5.7	993	19	AAW98240	H. pylori GHPO 741
13	112.5	5.7	1068	13	AAR20198	Sucrose phosphate
14	112.5	5.7	1068	18	AAW38266	Maize sucrose phos
15	112.5	5.7	1068	18	AAW09869	Sucrose phosphate
16	111	5.6	478	20	AAW92950	WO9905287 Seq ID 1
17	110	5.6	469	21	AAG31533	Arabidopsis thalia
18	110	5.6	503	21	AAG31532	Arabidopsis thalia
19	110	5.6	519	21	AAG31531	Arabidopsis thalia
20	110	5.6	542	21	AAG50592	Arabidopsis thalia
21	110	5.6	576	21	AAG50591	Arabidopsis thalia
22	110	5.6	592	21	AAG50590	Arabidopsis thalia
23	109	5.5	609	13	AAR23174	Mutant thermostabl
24	109	5.5	690	13	AAR23173	Mutant thermostabl
25	109	5.5	754	13	AAR23172	Mutant thermostabl
26	109	5.5	800	13	AAR23171	Mutant thermostabl
27	109	5.5	856	13	AAR23170	Mutant thermostabl
28	108.5	5.5	372	19	AAW79295	Neiserria meningit
29	108.5	5.5	1068	13	AAR27931	SPS protein. Zea
30	108	5.5	892	13	AAR23169	Mutant thermostabl
31	108	5.5	892	13	AAR23122	Taf DNA polymerase
32	107.5	5.5	1411	17	AAW02258	Nucleolar/endosoma
33	107	5.4	552	20	AAW96299	vgaB streptogramin
34	107	5.4	668	21	AAY32265	Catalpa speciosa s
35	106	5.4	1254	11	AAR07503	Merozite apical-en
36	106	5.4	1254	18	AAW24575	Merozoite apical-e
37	104.5	5.3	473	19	AAW64392	S. cerevisiae L947
38	104.5	5.3	1198	19	AAW64384	S. cerevisiae L947
39	104	5.3	403	18	AAW11218	Leishmania brazili
40	104	5.3	1634	21	AAY52023	M. jannaschii MJ08
41	104	5.3	1634	21	AAY51652	M. jannaschii MJ08
42	102.5	5.2	367	21	AAY93232	An Escherichia col
43	102.5	5.2	416	20	AAW92993	S. aureus ftsY pro
44	102	5.2	1123	21	AAY58277	Heliothis armigera
45	101.5	5.2	340	21	AAG14828	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAY33444
ID AAY33444 standard; Protein; 382 AA.
XX
AC AAY33444;
XX
DT 13-DEC-1999 (first entry)
XX
DE B. subtilis glycosyl transferase catalytic domain.
XX
KW Catalytic domain; glycosyl transferase; processive activity; detergent;
KW oligosaccharide glycolipid; 3-oligoglucosyl-1,2-diacylglycerol;
KW food emulsifier; polymer modifier; glycosyl transferase.
XX
OS Bacillus subtilis.
XX
PN DE19819958-A1.
XX
PD 30-SEP-1999.
XX
PF 05-MAY-1998; 98DE-1019958.
XX
PR 25-MAR-1998; 98DE-1013017.
XX
PA (GVSE-) GVS GES ERWERB & VERW LANDWIRTSCHAFTLICH.
PA (BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.
XX
PI Wolter FP, Jorasch P, Heinz E, Zaehringer U;
XX
DR WPI; 1999-552364/47.
DR N-PSDB; AAZ23386.
XX
PT New glycosyl transferase protein, useful for producing glycolipids -
XX

PS Example 1; Page 7; 8pp; German.
XX
CC This invention describes a novel protein (I) comprising identical or
CC different catalytic domains of glycosyl transferases and which has
CC processive activity (i.e. builds up oligosaccharide glycolipids by
CC successive addition of glucose units). (I) can be used to produce
CC 3-oligoglucosyl-1,2-diacylglycerols useful as food emulsifiers, polymer
CC modifiers or detergents. This sequence represents a Bacillus subtilis
CC glycosyl transferase catalytic domain which is used in the method of
CC the invention.
XX
SQ Sequence 382 AA;

Query Match 100.0%; Score 1970; DB 20; Length 382;
Best Local Similarity 100.0%; Pred. No. 1.6e-180;
Matches 382; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MNTNKRVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLK 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mntnkrvliltanygnghvqvaktlyeqcvrlgfqhvtvsnlyqesnpivsevtqylylk 60

Qy      61 SFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEY 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 sfsigkqfyrlfyygvdkiynkrkfniyfkmgnkrlgelvdehqpdiiintfpmivvpey 120

Qy     121 RRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKITGIPIR 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 rrrtgrviptfnvmtdfclhkiwvhenvdkyyvatdyvkeklleigthpsnvkitgipir 180

Qy     181 PQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVVVCGKN 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 pqfeesmpvgpiykkynlspnkkvllimagahgvlknvkelcenlvkddqvqvvvvcgkn 240

Qy     241 TALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGITLTEATAIGVPVILYK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 talkeslsaleaengdklkvlgyveridelfritdcmitkpggitlteataigvpvilyk 300

Qy     301 PVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDTLHRMKKNIKDLHLANSSEV 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 pvpgqekenanffedrgaaivvnrheeilesvtslladedtlhrmkknikdlhlanssev 360

Qy     361 ILEDILKESEMMTAKQKAKVLS 382
          ||||||||||||||||||||||
Db     361 iledilkesemmtakqkakvls 382
```

RESULT 2
AAB19039
ID AAB19039 standard; protein; 522 AA.
XX
AC AAB19039;
XX
DT 08-FEB-2001 (first entry)
XX
DE Amino acid sequence of monogalactosyl-diacylglycerol (MGDG) synthase.
XX
KW Monogalactosyl-diacylglycerol synthase; MGDG synthase; Toxoplasma;
KW apicomplex parasite; herbicide; antiparasitic; Plasmodium; Eimeria;
KW acquired immune deficiency syndrome; coccidiosis.
XX
OS Spinacia oleracea.
XX
PN WO200056919-A1.
XX
PD 28-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-FR00658.
XX
PR 19-MAR-1999; 99FR-0003434.
XX
PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI Marechal E, Block M, Joyard J, Douce R;
XX
DR WPI; 2000-602227/57.
XX
PT Use of monogalactosyl diacylglycerol synthase for identifying its
PT specific inhibitors, potentially useful as antiparasitic agents and
PT herbicides -
XX
PS Example 2; Fig 2; 33pp; French.
XX
CC The present sequence represents a monogalactosyl-diacylglycerol (MGDG)
CC synthase. MGDG is present in all plastids tested and is essential for
CC cell survival, but is not present in other membrane structures, or in
CC animal cells, and so represents a specific target. MGDG synthase cr a
CC plastid membrane isolated from plants is used to select and screen for
CC specific inhibitors of MGDG synthase. These inhibitors are suitable as
CC active agents against apicomplex parasites or as herbicides. The
CC inhibitors are used as antiparasitic agents, especially against
CC Plasmodium, Toxoplasma (particularly in patients with acquired immune
CC deficiency syndrome) or Eimeria (coccidiosis in cattle or poultry), and
CC as herbicides.
XX
SQ Sequence 522 AA;

Query Match 16.5%; Score 326; DB 21; Length 522;
Best Local Similarity 25.3%; Pred. No. 1.2e-22;
Matches 98; Conservative 80; Mismatches 162; Indels 48; Gaps 12;

```
Qy       5 KRVLILTANYGNGHVQVAKTLYEQCVRLGFQ-------HVTVSNLYQESN--PIVSEVTQ 55
          |||||| :: | ||   |     : ::  |          | |::|: |    |
Db     131 krvliimsdtggghrasa-----eaikaafneefgddyqvfvtdlwsehtpwpfnqlprs 185

Qy      56 YLYLKSFSIGKQFYRLFYYGVD-KIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM 114
          | :|         ::: |||   :: ::  |        : :   : ::|||||||:  |:
Db     186 ynflvkhg---plwkmmyygtsprvihqsnfaatsvfiarevarglmkyqpdiiisvhpl 242

Qy     115 IVVPEYRRRTGR----VIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHP 169
          :     |   ||     |    |:||    |  | |: | : |  :: | ::  : | |
Db     243 mqhvplrilrgrgllekivfttvvtdlstchptwfhklvtrcycpsnevakratkaglqp 302

Qy     170 SNVKITGIPIRPQFEESM-PVGPIYKKYNLSPNKKVLLIMAGAHGV--LKNVKELCENLV 226
          | :|: |:|:|| |  |: |   : |:  :  :   :|:| |  |:  ::       | :
Db     303 sqikvyglpvrpsfvrsvrpknelrkelgmdehlpavllmgggegmgpieataralgnal 362

Qy     227 KDDQV-----QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKP 281
            |  :     |::|:||:|  |   ||::: :    ::| |:| :|:|     ||:|||
Db     363 ydanlgeptgqllvicgrnkklagklssidwk--ipvqvkgfvtkieecmgacdciitka 420

Qy     282 GGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDT 341
          |  |: ||   |:|:||   : ||| |  :  : |    :   :|| ::|:     :
Db     421 gpgtiaeamirglpiilndyiagqeagnvpyviengigkylkspkeiaktvsqwfgpk-- 478

Qy     342 LHRMKKNIKDLHLANSSEVILEDILKES 369
                    ||  ::: :: || :
Db     479 ------------anelqimsqnalkha 493
```

RESULT 3
AAG42414
ID AAG42414 standard; Protein; 492 AA.
XX
AC AAG42414;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52894.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.

XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

```
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             15.7%;  Score 309.5;  DB 21;  Length 492;
  Best Local Similarity   24.2%;  Pred. No. 4.1e-21;
  Matches  102;  Conservative   91;  Mismatches  171;  Indels   57;  Gaps   17;

Qy       5 KRVLILTANYGNGHVQVAKTL----------------YEQCVRLGFQ---HVTVSNLYQ 44
           |:|||| :: | ||   |: :                 : :  ::||:   || :::|:
Db      83 kkvlilmsdtggghrasaeairaafnqefgdeyqsgleflelekigfrclvhvfitdlwt 142

Qy      45 ESN--PIVSEVTQYLYLKSFSIGKQFYRLFYYGVD-KIYNKRKFNIYFKMGNKRLGELVD 101
           :   |       | :|         ::: |||   :| ::  |        : :.: :
Db     143 dhtpwpfnqlprsynflvkhg---tlwkmtyygtsprivhqsnfaatstfiareiaqglm 199

Qy     102 EHQPDIIINTFPMI--VVPEYRRRTG--RVIPTFNVMTDF-CLHKIWVHENVDKYYVATD 156
           ::||||||:  |::  |     |  |  : |    |:||    |  | |: | : |  :
Db     200 kyqpdiiisvhplmqhvplrvlrskgllkkivfttvitdlstchptwfhklvtrcycpst 259

Qy     157 YVKEKLLEIGTHPSNVKITGIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGAHG 213
             | ::  : |   | :|: |:|:|| | :  || |   : ::  :  |   :|:| |  |
Db     260 evakraqkagletsqikvyglpvrpsfvk--pvrpkvelrrelgmdenlpavllmgggeg 317

Qy     214 V--LKNVKELCENLVKDDQV-----QVVVVCGKNTALKESLSALEAENGDKLKVLGYVER 266
           :  ::       : : | :     ||:::||:|  |:  ||:|: :    ::| |:: :
Db     318 mgpieataraladalydknlgeavgqvliicgrnkklqsklssldwk--ipvqvkgfitk 375

Qy     267 IDELFRITDCMITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHE 326
           ::|     ||:||| |  |: ||   |:|:||   : |||  |  :  : |      :
Db     376 meecmgacdciitkagpgtiaeamirglpiilngyiagqeagnvpyvvengc----gkfs 431

Qy     327 EILESVTSLLAD-----EDTLHRMKKNIKDLHLANSSEVILEDILKESEMMTAKQKAKVL 381
           :  : :: ::||        |  | :|   | ||   | : : :    |::  |     |
Db     432 kspkeiskivadwfgpaskeleimsqna--lrla-kpeavfkivhdmhelvrkknslpql 488

Qy     382 S 382
           |
Db     489 s 489


RESULT   4
AAG42413
ID   AAG42413 standard; Protein; 551 AA.
XX
AC   AAG42413;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 52893.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
```

PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 15.7%; Score 309.5; DB 21; Length 551;
Best Local Similarity 24.2%; Pred. No. 4.8e-21;
Matches 102; Conservative 91; Mismatches 171; Indels 57; Gaps 17;

```
Qy      5 KRVLILTANYGNGHVQVAKTL----------------YEQCVRLGFQ---HVTVSNLYQ 44
          |:|||| :: | ||   |: :                : :  ::||:   || :::|:
Db    142 kkvlilmsdtggghrasaeairaafnqefgdeyqsgleflelekigfrclvhvfitdlwt 201

Qy     45 ESN--PIVSEVTQYLYLKSFSIGKQFYRLFYYGVD-KIYNKRKFNIYFKMGNKRLGELVD 101
           :    |       | :|         ::: |||   :| ::  |       : : : :
```

```
Db    202 dhtpwpfnqlprsynflvkhg---tlwkmtyygtsprivhqsnfaatstfiareiaqglm 258

Qy    102 EHQPDIIINTFPMI--VVPEYRRRTG--RVIPTFNVMTDF-CLHKIWVHENVDKYYVATD 156
          ::||||||:  |::  |     |  |  : |    |:||    |  | |: | : |  :
Db    259 kyqpdiiisvhplmqhvplrvlrskgllkkivfttvitdlstchptwfhklvtrcycpst 318

Qy    157 YVKEKLLEIGTHPSNVKITGIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGAHG 213
           | ::  : |   | :|: |:|:|| | :  || |   : ::  :  |   :|:| |  |
Db    319 evakraqkagletsqikvyglpvrpsfvk--pvrpkvelrrelgmdenlpavllmgggeg 376

Qy    214 V--LKNVKELCENLVKDDQV-----QVVVVCGKNTALKESLSALEAENGDKLKVLGYVER 266
          :  ::       : : |  :     ||:|:||:|  |:  ||:|: :    ::| |:: :
Db    377 mgpieataraladalydknlgeavgqvlticgrnkklqsklssldwk--ipvqvkgfitk 434

Qy    267 IDELFRITDCMITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHE 326
          ::|    ||:||| |  |: ||   |:|:||   : |||  |  :  : |      :
Db    435 meecmgacdciitkagpgtiaeamirglpiilngyiagqeagnvpyvvengc----gkfs 490

Qy    327 EILESVTSLLAD-----EDTLHRMKKNIKDLHLANSSEVILEDILKESEMMTAKQKAKVL 381
          :  : :: ::||        |  | :|   | ||   | : : :    |:: |     |
Db    491 kspkeiskivadwfgpaskeleimsqna--lrla-kpeavfkivhdmhelvrkknslpql 547

Qy    382 S 382
          |
Db    548 s 548
```

RESULT 5
AAW41148
ID AAW41148 standard; Protein; 422 AA.
XX
AC AAW41148;
XX
DT 05-MAY-1998 (first entry)
XX
DE Monogalactosyldiacylglycerol synthetase sequence.
XX
KW Monogalactosyldiacylglycerol synthetase; MGDG; lipid production;
KW cucumber.
XX
OS Cucumis sativus.
XX
PN JP10014579-A.
XX
PD 20-JAN-1998.
XX
PF 02-JUL-1996; 96JP-0172337.
XX
PR 02-JUL-1996; 96JP-0172337.
XX
PA (KIRI) KIRIN BREWERY KK.
XX
DR WPI; 1998-138241/13.
DR N-PSDB; AAV12734.
XX
PT New mono:galactosyl-di:acyl-glycerol synthase - useful for, e.g.
PT increasing lipid production in transformed host
XX
PS Claim 1; Page 8; 13pp; Japanese.
XX
CC This sequence represents the monogalactosyldiacylglycerol synthase (MGDG)
CC of the invention. This sequence was isolated from cucumber. MGDG may be
CC used to increase lipid production in an organism.
XX
SQ Sequence 422 AA;

Query Match 15.5%; Score 305.5; DB 19; Length 422;
Best Local Similarity 25.7%; Pred. No. 7.9e-21;
Matches 100; Conservative 78; Mismatches 178; Indels 33; Gaps 13;

```
Qy      5 KRVLILTANYGNGH---VQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLKS 61
            |||||| :: | ||    :  |  : :     :| | :::|: :  |         |

Db     33 krvlilmsdtggghrasaeaikaafneefgnnyq-vfitdlwtdhtpwpfnqlprsynfl 91

Qy     62 FSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM-----IV 116
             |  :   :|    |: ::  |        : : : : :::|||||:  |:     |
Db     92 vkhgtlwkmtyyvtapkvihqsnfaatstfiarevakglmkyrpdiiisvhplmqhvpir 151

Qy    117 VPEYRRRTGRVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKIT 175
          :   :    ::: |  |:||    |  | |: | : |  :  | :: |: |  || :|:
Db    152 ilrskgllnkivft-tvvtdlstchptwfhklvtrcycpstevakralkaglqpsklkvf 210

Qy    176 GIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGAHG---VLKNVKELCENLVKDD 229
          |:|:|| | :  |: |   : |:  :  |   :|:| |  |   :    | | : |  ::
Db    211 glpvrpsfvk--pirpkielrkelgmdenlpavllmgggegmgpieatakalskalyden 268

Qy    230 Q----VQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGIT 285
              ||:|:|| |  |   | ::: :    ::| |:| :::|     ||:||| |  |
Db    269 hgepigqvlvicghnkklagrlrsidwk--vpvqvkgfvtkmeecmgacdciitkagpgt 326

Qy    286 LTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADE-DTLHR 344
          : ||   |:|:||   : |||  |  :  : |         :||   |      : | |
Db    327 iaeamirglpiilndyiagqeagnvpyvvengcgkfskspkeianivakwfgpkadelli 386

Qy    345 MKKNIKDLHLANSSEVI-----LEDILKE 368
          | :|   | ||    |      | :::|:
Db    387 msqna--lrlarpdavfkivhdlhelvkq 413
```

RESULT 6
AAB19040
ID AAB19040 standard; protein; 525 AA.
XX
AC AAB19040;
XX
DT 08-FEB-2001 (first entry)
XX
DE Amino acid sequence of monogalactosyl-diacylglycerol (MGDG) synthase.
XX
KW Monogalactosyl-diacylglycerol synthase; MGDG synthase; Toxoplasma;
KW apicomplex parasite; herbicide; antiparasitic; Plasmodium; Eimeria;
KW acquired immune deficiency syndrome; coccidiosis.
XX
OS Cucumis sativus.
XX
PN WO200056919-A1.
XX
PD 28-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-FR00658.
XX
PR 19-MAR-1999; 99FR-0003434.
XX
PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI Marechal E, Block M, Joyard J, Douce R;
XX
DR WPI; 2000-602227/57.
XX
PT Use of monogalactosyl diacylglycerol synthase for identifying its
PT specific inhibitors, potentially useful as antiparasitic agents and
PT herbicides -
XX
PS Example 2; Fig 2; 33pp; French.
XX
CC The present sequence represents a monogalactosyl-diacylglycerol (MGDG)
CC synthase. MGDG is present in all plastids tested and is essential for
CC cell survival, but is not present in other membrane structures, or in
CC animal cells, and so represents a specific target. MGDG synthase or a
CC plastid membrane isolated from plants is used to select and screen for
CC specific inhibitors of MGDG synthase. These inhibitors are suitable as
CC active agents against apicomplex parasites or as herbicides. The
CC inhibitors are used as antiparasitic agents, especially against
CC Plasmodium, Toxoplasma (particularly in patients with acquired immune
CC deficiency syndrome) or Eimeria (coccidiosis in cattle or poultry), and

```
CC   as herbicides.
XX
SQ   Sequence   525 AA;


  Query Match            15.4%;  Score 303.5;  DB 21;  Length 525;
  Best Local Similarity   25.7%;  Pred. No. 1.7e-20;
  Matches  100;  Conservative   77;  Mismatches  179;  Indels   33;  Gaps   13;

Qy       5 KRVLILTANYGNGH---VQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLKS 61
           |||||| :: | ||    :  |  : :     :| | :::|: :  |          |
Db     136 krvlilmsdtggghrasaeaikaafneefgnnyq-vfitdlwtdhtpwpfnqlprsynfl 194

Qy      62 FSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM-----IV 116
             |  :   :|    |: ::  |        : : : : :::|||||:  |:     |
Db     195 vkhgtlwkmtyyvtapkvihqsnfaatstfiarevakglmkyrpdiiisvhplmqhvpir 254

Qy     117 VPEYRRRTGRVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKIT 175
           :   :    ::: |  |:||    |  | |: | : |  :  | :: |  |  || :|:
Db     255 ilrskgllnkivft-tvvtdlstchptwfhklvtrcycpstevakraltaglqpsklkvf 313

Qy     176 GIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGAHG---VLKNVKELCENLVKDD 229
           |:|:|| | :  |: |   : |:  :  |   :|:| |  |   :    | | : |  ::
Db     314 glpvrpsfvk--pirpkielrkelgmdenlpavllmgggegmgpieatakalskalyden 371

Qy     230 Q----VQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGIT 285
             ||:|:|| |  |   | ::: :    ::| |:| :::|     ||:||| |  |
Db     372 hgepigqvlvicghnkklagrlrsidwk--vpvqvkgfvtkmeecmgacdciitkagpgt 429

Qy     286 LTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADE-DTLHR 344
           : ||   |:|:||   : |||  |  :  : |        :||   |      : | |
Db     430 iaeamirglpiilndyiagqeagnvpyvvengcgkfskspkeianivakwfgpkadelli 489

Qy     345 MKKNIKDLHLANSSEVI-----LEDILKE 368
           | :|   | ||    |      | :::|:
Db     490 msqna--lrlarpdavfkivhdlhelvkq 516


RESULT   7
AAB19041
ID   AAB19041 standard; protein; 533 AA.
XX
AC   AAB19041;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Amino acid sequence of monogalactosyl-diacylglycerol (MGDG) synthase.
XX
KW   Monogalactosyl-diacylglycerol synthase; MGDG synthase; Toxoplasma;
KW   apicomplex parasite; herbicide; antiparasitic; Plasmodium; Eimeria;
KW   acquired immune deficiency syndrome; coccidiosis.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200056919-A1.
XX
PD   28-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-FR00658.
XX
PR   19-MAR-1999;   99FR-0003434.
XX
PA   (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI   Marechal E,  Block M,  Joyard J,  Douce R;
XX
DR   WPI; 2000-602227/57.
XX
PT   Use of monogalactosyl diacylglycerol synthase for identifying its
PT   specific inhibitors, potentially useful as antiparasitic agents and
PT   herbicides -
XX
PS   Example 2; Fig 2; 33pp; French.
XX
CC   The present sequence represents a monogalactosyl-diacylglycerol (MGDG)
CC   synthase. MGDG is present in all plastids tested and is essential for
CC   cell survival, but is not present in other membrane structures, or in
CC   animal cells, and so represents a specific target. MGDG synthase or a
CC   plastid membrane isolated from plants is used to select and screen for
CC   specific inhibitors of MGDG synthase. These inhibitors are suitable as
CC   active agents against apicomplex parasites or as herbicides. The
CC   inhibitors are used as antiparasitic agents, especially against
CC   Plasmodium, Toxoplasma (particularly in patients with acquired immune
CC   deficiency syndrome) or Eimeria (coccidiosis in cattle or poultry), and
CC   as herbicides.
XX
SQ   Sequence   533 AA;


  Query Match            15.3%;  Score 301.5;  DB 21;  Length 533;
  Best Local Similarity   24.8%;  Pred. No. 2.7e-20;
  Matches  101;  Conservative   86;  Mismatches  172;  Indels   49;  Gaps   17;

Qy       5 KRVLILTANYGNGHVQVAKTLYEQCVRLGFQ-------HVTVSNLYQESN--PIVSEVTQ 55
           |:|||| :: | ||   |     : :|  |         | :::|: :    |
Db     142 kkvlilmsdtggghrasa-----eairaafnqefgdeyqvfitdlwtdhtpwpfnqlprs 196

Qy      56 YLYLKSFSIGKQFYRLFYYGVD-KIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM 114
           | :|          ::: |||   :| ::  |        : : : : ::||||||:  |:
Db     197 ynflvkhg---tlwkmtyygtsprivhqsnfaatstfiareiaqglmkyqpdiiisvhpl 253

Qy     115 I--VVPEYRRRTG--RVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHP 169
           :  |     |  | : |    |:||    |  | |: | : |  :  | ::  : |
Db     254 mqhvplrvlrskgllkkivfttvitdlstchptwfhklvtrcycpstevakraqkaglet 313

Qy     170 SNVKITGIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGAHGV--LKNVKELCEN 224
           | :|: |:|:|| | :  || |   : ::  :  |   :|:| |  |:  ::       :
Db     314 sqikvyglpvrpsfvk--pvrpkvelrrelgmdenlpavllmgggegmgpieataralad 371

Qy     225 LVKDDQV-----QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMIT 279
           : |  :      ||:::||:|  |:  ||:|: :    ::| |:: :::|     ||:||
Db     372 alydknlgeavgqvliicgrnkklqsklssldwk--ipvqvkgfitkmeecmgacdciit 429

Qy     280 KPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLAD- 338
           | |  |: ||   |:|:||   : |||  |  :  : |      :  :  : :: ::||
Db     430 kagpgtiaeamirglpiilngyiagqeagnvpyvvengc----gkfskspkeiskivadw 485

Qy     339 ----EDTLHRMKKNIKDLHLANSSEVILEDILKESEMMTAKQKAKVLS 382
                |  | :|   | ||   | : : :    |::  |     ||
Db     486 fgpaskeleimsqna--lrla-kpeavfkivhdmhelvrkknslpqls 530


RESULT   8
AAB19042
ID   AAB19042 standard; protein; 468 AA.
XX
AC   AAB19042;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Amino acid sequence of monogalactosyl-diacylglycerol (MGDG) synthase.
XX
KW   Monogalactosyl-diacylglycerol synthase; MGDG synthase; Toxoplasma;
KW   apicomplex parasite; herbicide; antiparasitic; Plasmodium; Eimeria;
KW   acquired immune deficiency syndrome; coccidiosis.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200056919-A1.
XX
PD   28-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-FR00658.
XX
PR   19-MAR-1999;   99FR-0003434.
XX
```

PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI Marechal E, Block M, Joyard J, Douce R;
XX
DR WPI; 2000-602227/57.
XX
PT Use of monogalactosyl diacylglycerol synthase for identifying its
PT specific inhibitors, potentially useful as antiparasitic agents and
PT herbicides -
XX
PS Example 2; Fig 2; 33pp; French.
XX
CC The present sequence represents a monogalactosyl-diacylglycerol (MGDG)
CC synthase. MGDG is present in all plastids tested and is essential for
CC cell survival, but is not present in other membrane structures, or in
CC animal cells, and so represents a specific target. MGDG synthase or a
CC plastid membrane isolated from plants is used to select and screen for
CC specific inhibitors of MGDG synthase. These inhibitors are suitable as
CC active agents against apicomplex parasites or as herbicides. The
CC inhibitors are used as antiparasitic agents, especially against
CC Plasmodium, Toxoplasma (particularly in patients with acquired immune
CC deficiency syndrome) or Eimeria (coccidiosis in cattle or poultry), and
CC as herbicides.
XX
SQ Sequence 468 AA;

Query Match 14.9%; Score 293; DB 21; Length 468;
Best Local Similarity 25.1%; Pred. No. 1.5e-19;
Matches 98; Conservative 86; Mismatches 172; Indels 34; Gaps 14;

```
Qy       5 KRVLILTANYGNGHVQVAKTLYEQC-VRLGFQH-VTVSNLYQESN--PIVSEVTQYLYLK 60
           | |||| :: | ||   |: : :   :  | :: | | ::::|    |:      | ::
Db      69 knvlilmsdtggghrasaeairdafkiefgdkyrvivkdvwkeytgwplndmersykfmv 128

Qy      61 SFSIGKQFYRLFYYG-----VDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM- 114
             | ::: ::      :   |       | |     |     |::|:|||:  |:
Db     129 kh---vqlwkvafhstspkwihscylaaiaayyakeveagl----meykpeiiisvhplm 181

Qy     115 ----IVVPEYRRRTGRVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHP 169
             : | :::    ||:    |:||    |  | |  |:: |  :  | :: |  |
Db     182 qhiplwvlkwqelqkrvl-fvtvitdlntchptwfhpgvnrcycpsqevakralfdglde 240

Qy     170 SNVKITGIPIRPQFEESMPV-GPIYKKYNLSPNKKVLLIMAGAHGV--LKNVKELCENLV 226
           | |:: |:|:|| |  :: |   : |:  :  : : :|:| |  |:  :|   :  |  :
Db     241 sqvrvfglpvrpsfaravlvkddlrkelemdqdlravllmgggegmgpvketakaleefl 300

Qy     227 KDDQ-----VQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKP 281
           | :      |:||:||:|  |  :| |:: :    :|| |:  ::::     ||:|||
Db     301 ydkenrkpigqmvvicgrnkklasaleaidwk--ipvkvrgfetqmekwmgacdciitka 358

Qy     282 GGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADE-D 340
           |  |: |:    :|:||   :|||||  |  :  : || :     :|    |    : : |
Db     359 gpgtiaeslirslpiilndyipgqekgnvpyvvengagvftrspketarivgewfstktd 418

Qy     341 TLHRMKKNIKDLHLANSSEVILEDILKESE 370
           | :   | : |    :   |::|| : ||
Db     419 eleqtsdnarklaqpeavfdivkdidelse 448
```

RESULT 9
AAG42415
ID AAG42415 standard; Protein; 404 AA.
XX
AC AAG42415;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52895.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD . 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.6%; Score 287.5; DB 21; Length 404;
Best Local Similarity 23.6%; Pred. No. 3.9e-19;
Matches 97; Conservative 88; Mismatches 169; Indels 57; Gaps 17;

```
Qy      15 GNGHVQVAKTL----------------YEQCVRLGFQ---HVTVSNLYQESN--PIVSE 52
Db       5 ggghrasaeairaafnqefgdeyqsglefllelekigfrclvhvfitdlwtdhtpwpfnql 64

Qy      53 VTQYLYLKSFSIGKQFYRLFYYGVD-KIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINT 111
Db      65 prsynflvkhg---tlwkmtyygtsprivhqsnfaatstfiareiaqglmkyqpdiiisv 121

Qy     112 FPMI--VVPEYRRRTG--RVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIG 166
Db     122 hplmqhvplrvlrskgllkkivfttvitdlstchptwfhklvtrcycpstevakraqkag 181

Qy     167 THPSNVKITGIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGAHGV--LKNVKEL 221
Db     182 letsqikvyglpvrpsfvk--pvrpkvelrrelgmdenlpavllmgggegmgpieatara 239

Qy     222 CENLVKDDQV-----QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDC 276
Db     240 ladalydknlgeavgqvliicgrnkklqsklssldwk--ipvqvkgfitkmeecmgacdc 297

Qy     277 MITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLL 336
Db     298 iitkagpgtiaeamirglpiilngyiagqeagnvpyvvengc----gkfskspkeiskiv 353

Qy     337 AD-----EDTLHRMKKNIKDLHLANSSEVILEDILKESEMMTAKQKAKVLS 382
```

```
Db    354 adwfgpaskeleimsqna--lrla-kpeavfkivhdmhelvrkknslpqls 401


RESULT 10
AAW92952
ID   AAW92952 standard; Protein; 468 AA.
XX
AC   AAW92952;
XX
DT   14-MAY-1999  (first entry)
XX
DE   WO9905287 Seq ID 12.
XX
KW   Plant; flavonoid 5-transglycosylation activity; 5TG; variety; colour.
XX
OS   Petunia hybrida.
XX
PN   WO9905287-A1.
XX
PD   04-FEB-1999.
XX
PF   16-JUL-1998;   98WO-JP03199.
XX
PR   25-JUL-1997;   97JP-0200571.
XX
PA   (SUNR ) SUNTORY LTD.
XX
PI   Gong Z,  Kusumi T,  Mizutani M,  Saito K,  Tanaka Y;
PI   Yamazaki M;
XX
DR   WPI; 1999-142940/12.
DR   N-PSDB; AAX02831.
XX
PT   Gene encoding a plant flavonoid 5-transglycosylation enzyme - can be
PT   used to transform plants for improvement of plant coloration
XX
PS   Claim 2; Page 83-85; 89pp; Japanese.
XX
CC   This invention describes the isolation of plant proteins which have
CC   flavonoid 5-transglycosylation (5TG) activity. Such proteins can be used
CC   in the production of plant varieties with improved colour, for example in
CC   the production of flower varieties with desired coloration and cut
CC   flowers harvested from them.
XX
SQ   Sequence   468 AA;


  Query Match             5.8%;  Score 114.5;  DB 20;  Length 468;
  Best Local Similarity   20.4%;  Pred. No. 0.019;
  Matches   76;  Conservative   59;  Mismatches  130;  Indels  107;  Gaps   19;

Qy      42 LYQESNPIVSEVTQYLYLKSFSIGKQ------FYRLFYYGVDKIY----NKRKFNI---- 87
Db     110 lysiflpwaaevarevhipsallwsqpatildiyyfnfhgyekamanesndpnwsiqlpg 169

Qy      88 --------------YFKMGNKRLG-----ELVD----EHQPDIIINTFPMIVVPEYRRR 123
Db     170 lplletrdlpsfllpygakgslrvalppfkelidtldaettpkilvntfdel-epe---- 224

Qy     124 TGRVIPTFNVMTDFCLHKIWVHENVDKVVATDYVKEKLLEIGTHPSNVKITGIPIRPQF 183
Db     225 ------alnaiegy-----------kfygigplipsaflg-gndpldasfgg----dlf 261

Qy     184 EESMPVGPIYKKY-NLSPNKKVLLIMAGA--HGVLKNVKELCENLVKDDQVQVVVVCGKN 240
Db     262 qnsnd----ymewlnskpnssvvyisfgslmnpsisqmeeiskglidigrpflwvi---- 313

Qy     241 TALKESLSALEAENGDKLKVLGYVERIDELFRIT-------------DCMITKPGGITL 286
Db     314 ---kenekgkeeen----kklgcieelekigkivpwcsqlevlkhpslgcfvshcgwnsa 366

Qy     287 TEATAIGVPVILYKPVPGQEKENANFFED---RGAAIVVN-----RHEEILESVTSLLAD 338
```

```
Db    367 leslacgvpvvaf-pqwtdqmtnakqvedvwksgvrvrinedgvveseeikrcielvmdg 425

Qy    339 EDTLHRMKKNIK 350
Db    426 gekgeelrknak 437


RESULT 11
AAY74401
ID   AAY74401 standard; Protein; 355 AA.
XX
AC   AAY74401;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 087 protein sequence SEQ ID NO:278.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53163.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 277-278; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   355 AA;


  Query Match             5.8%;  Score 113.5;  DB 21;  Length 355;
  Best Local Similarity   20.3%;  Pred. No. 0.016;
```

```
  Matches   80;  Conservative   71;  Mismatches  158;  Indels   85;  Gaps   18;

Qy       6 RVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSE--VTQYLYLKSFS 63
           :  :| |    ||:  |  : :  :|:  ||    ::  |   : |  | ||
Db       4 ktfmlmaggtgghifpalavads-lrvrghhv----iwlgskdsmeerivpqy------- 51

Qy      64 IGKQFYRLFYYGVDKIYNKRKFNIYFKMGN--KRLGELVDEHQPDIIIN-----TFP--- 113
             | :   |   |:     |||  : | :    :    :: :|: : :|      |||
Db      52 -girletlaikgirgngikrklmlpftlyktvreaqriirkhrvecvigfggfvtfpggl 110

Qy     114 -------MIVVPEYRRRTG---RVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLL 163
                  ||: |     |   | :  :     :   | : ||
Db     111 aakllgvpiviheqnavaglsnrhlsrwakrvlyafpkafsheg---------------- 154

Qy     164 EIGTHPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLLI--MAGAHGVLKNVKEL 221
                 : | |:|     ::|| |  :        |:|::    ||  : | | :
Db     155 ---------glvgnpvradi-snlpv-paerfqgregrlkilvvggslgadvlnktvpqa 203

Qy     222 CENLVKDDQVQVVVVCGKNTALKESLSALEAEN---GDKLKVLGYVERIDELFRITDCMI 278
           | :: : |:    |:|      |  |:|:    | | : : ::  :   :|  | :|
Db     204 lallpeevrpqmyhqsgrn-----klgnlqadydalgvkaecvefitdmvsayrdadlvi 258

Qy     279 TKPGGITLTEATAIGVPVILYKPVP----GQEKENANFFEDRGAAIVVNRHEEILESVTS 334
           : | :|: | || |:  :|  | |      :  || |     | ::: : :   | :
Db     259 cragaltiaeltaaglgallv-pyphavddhqtanarfmvqaeagllpqtqltaeklae 317

Qy     335 LLADEDTLHRMK-----KNIKDLHLANSSEVILE 363
           :|    :|:| |     :| : | | :|:: : |
Db     318 ilg---slnrekclkwaenartlalphsaddvae 348


RESULT  12
AAW98240
ID   AAW98240 standard; Protein; 993 AA.
XX
AC   AAW98240;
XX
DT   31-MAR-1999  (first entry)
XX
DE   H. pylori GHPO 741 protein.
XX
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease.
XX
OS   Helicobacter pylori.
XX
PN   WO9843478-A1.
XX
PD   08-OCT-1998.
XX
PF   01-APR-1998;   98WO-US06371.
XX
PR   29-JUL-1997;   97US-0902615.
PR   01-APR-1997;   97US-0833457.
PR   24-JUN-1997;   97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Kleanthous H,  Miller C,  Oomen RP,  Tomb J;
XX
DR   WPI; 1998-542293/46.
DR   N-PSDB; AAX13959.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
XX
PS   Claim 8; Page 200-204; 2054pp; English.
XX
CC   This sequence represents a Helicobacter pylori GHPO protein of the
CC   invention. The polypeptides can be used for preventing or treating
CC   Helicobacter infections, and gastroduodenal diseases associated with
CC   these infections, including acute, chronic, and atrophic gastritis, and
CC   peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC   used for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
XX
SQ   Sequence   993 AA;


  Query Match             5.7%;  Score 112.5;  DB 19;  Length 993;
  Best Local Similarity   20.5%;  Pred. No. 0.091;
  Matches   95;  Conservative   70;  Mismatches  154;  Indels  145;  Gaps   21;

Qy       6 RVLILTANYGNGHVQVAKTLYEQCVRL--------------GFQHVTVSNLYQESNPIV 50
             |:::   :  :|      | : || : :                | : | |: |::    |
Db     101 rliilwekdtksgekngikdikeqsifireiplmtertsfiingvervvvvnqlhrspgvif 160

Qy      51 SEVTQYLYLKSFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMG---------NKR------ 95
            |         | |: |    | | |  :|   |   :||:           |||
Db     161 keee-----sstslnk----liytg--qiipdrgswlyfeydskdvlyarinkrrkvpvt 209

Qy      96 -LGELVDEHQPDIIINTFPMIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKY--- 151
            |   :|  : |||   :|:: |                            :|| |||
Db     210 ilframdyqkqdiikmfyplvkvr-------------------------yen-dkylip 242

Qy     152 YVATDYVKEKLLEIGTHPSNV------KITGIPIRPQFEESM-----PVGPIYKKYNLSP 200
           : : |  :    ::       |       |:|   |:   |  :     |:  :  ::   |
Db     243 fasldanqrmefdlkdpqgkvillagkkltsrkikelkenhlewveypmdillnrhlaep 302

Qy     201 NKKVLLIMAGAHGVLKNVKELCEN---LVKDDQVQVVVV-----CGKNTALKESLSALEA 252
               :| |   :|  : :| :|    : |  ||  |:       | : :: :| ||
Db     303 ------vmvgkevlldmltqldknklekihdlgvqefviindlalghdasiiqsfsa--- 353

Qy     253 ENGDKLKVLGYVERIDE---LFRITDCMITKPGGITLTEATAIGVPVILYKP------VP 303
             : : ||:|   |:||:   |  |    : |||    ||     |  : : |      :
Db     354 -dseslkllkqtekiddenalaairihkvmkpgdpvttevakqfvkklffdperydltmv 412

Qy     304 GQEKENANFFEDRGAAIVVNRHEEI-------------------------LESVTSLL 336
           |: | |           |    ||:|                        : :|  ||
Db     413 grmkmnhklglhvpdyittlthediittvkylmkiknnqgkiddrdhlgnrriravgell 472

Qy     337 ADEDTLH----RMKKNIKD--LHLANSSEVILEDILKESEMMTA 374
           |:|  ||    :|:| |||    :: : : ::   |  |:|:|:
Db     473 ane--lhsglvkmqktikdkllttmsgafdslmphdlvnskmits 514


RESULT  13
AAR20198
ID   AAR20198 standard; Protein; 1068 AA.
XX
AC   AAR20198;
XX
DT   14-APR-1992  (first entry)
XX
DE   Sucrose phosphate synthase from corn.
XX
KW   SPS; fructose-6-phosphate; UDP-glucose; sucrose regulation;
KW   carbon partitioning; plants.
XX
OS   Zea mays L. cv Pioneer 3184.
XX
FH   Key             Location/Qualifiers
FT   Peptide         71..74
FT                   /label= A8
FT                   /note=" tryptic "
FT   Peptide         206..212
FT                   /label= B4
FT                   /note=" tryptic "
FT   Peptide         471..481
FT                   /label= B11
FT                   /note=" tryptic "
FT   Peptide         872..884
FT                   /label= 4K
FT                   /note=" tryptic "
```

```
FT   Peptide          885..891
FT                    /label= 12N
FT                    /note=" tryptic "
XX
PN   EP466995-A.
XX
PD   22-JAN-1992.
XX
PF   20-JUL-1990;   90EP-0402084.
XX
PR   20-JUL-1990;   90EP-0402084.
XX
PA   (ROUS ) ROUSSEL UCLAF.
XX
PI   Van Assche C,  Lando D,  Bruneau JM,  Voelker TA,  Gervais M;
XX
DR   WPI; 1992-025895/04.
DR   N-PSDB; AAQ20664.
XX
PT   Sucrose phosphate synthase (SPS) from corn - its prepn., DNA
PT   encoding it, and hybridomas and monoclonal antibodies specific for
PT   it.
XX
PS   Claim 5; Fig 7; 40pp; English.
XX
CC   The sequence, which is that of sucrose phosphate synthase (SPS),
CC   was deduced from a cDNA sequence obtd. from three clones isolated
CC   from a cDNA library prepd. from RNA isolated from corn leaves. The
CC   tryptic peptides (see features) were used to design degenerative
CC   oligonucleotide primers and probes (see AAQ21987-93) for the prepn.
CC   by PCR of a probe for the SPS gene. The gene can be used to prepare
CC   a vector for expression of recombinant SPS, which is a key enzyme
CC   in sucrose regulation and carbon partitioning between starch and
CC   sucrose in photosynthesising plants. (Plants transformed with the
CC   DNA have an increased flow of sucrose to growing tissues and give
CC   increased yields). The recombinant SPS can be used to raise mono-
CC   clonal antibodies which can then be used to purify the protein by
CC   affinity chromatography.
XX
SQ   Sequence   1068 AA;
```

Query Match 5.7%; Score 112.5; DB 13; Length 1068;
Best Local Similarity 20.9%; Pred. No. 0.1;
Matches 56; Conservative 56; Mismatches 87; Indels 69; Gaps 14;

```
Qy     125 GRVIPTFNVM---TDFCLHKIWVHENVDKYYVATDYVKEKLLEI-GTHPSNVKITGIPIR 180
           || :|   |:    ||    : |||::|        ||: :: : |  | ::
Db     427 grymprmvvippgmdf--snvvvhedid----gdgdvkddivglegaspksmp------- 473

Qy     181 PQFEESM-----PVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVV 235
           | : | |     |  |:    :    || :  :  | |  : ::||          : :
Db     474 piwaevmrfltnphkpmilalsrpdpkknittlvkafgecrplrel---------anltl 524

Qy     236 VCGKNTALKESLSALEAENGDKL-KVLGYVERID----------------ELFRITDCM 277
           : |     :: :  : | |   |  ||  ::: |                 |::|:   |
Db     525 imgn----rddiddmsagnasvlttvlklidkydlygsvafpkhhnqadvpeiyrlaakm 580

Qy     278 --------ITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEE-- 327
                   : :| |:|| || | |:|::  |     :  ||          ::|: |::
Db     581 kgvfinpalvepfgltlieaaahglpivatknggpvditna-----lnngllvdphdqna 635

Qy     328 ILESVTSLLADEDTLHRMKKN-IKDLHL 354
           | :::  |:||::     ::| ::::||
Db     636 iadallklvadknlwqecrrnglrnihl 663
```

```
RESULT  14
AAW38266
ID   AAW38266 standard; Protein; 1068 AA.
XX
AC   AAW38266;
XX
```

```
DT   11-MAY-1998  (first entry)
XX
DE   Maize sucrose phosphate synthase.
XX
KW   Sucrose phosphate synthase; SPS; transgenic plant; maize; corn.
XX
OS   Zea mays L. cv. Pioneer 3184.
XX
FH   Key             Location/Qualifiers
FT   Peptide         71..74
FT                   /label= A8
FT   Peptide         206..212
FT                   /label= B4
FT   Peptide         471..481
FT                   /label= B11
FT   Peptide         872..885
FT                   /label= 4K
FT   Peptide         886..892
FT                   /label= 12N
XX
PN   EP807685-A2.
XX
PD   19-NOV-1997.
XX
PF   20-JUL-1990;   90EP-0402084.
XX
PR   20-JUL-1990;   90EP-0402084.
PR   20-JUL-1990;   90EP-0201062.
XX
PA   (ROUS ) ROUSSEL-UCLAF.
XX
PI   Bruneau J,  Gervais M,  Lando D,  Van Assche C,  Voelker TA;
XX
DR   WPI; 1997-552360/51.
DR   N-PSDB; AAT95847.
XX
PT   DNA encoding sucrose phosphate synthase - useful for producing
PT   transgenic plants
XX
PS   Claim 5; Fig 7; 38pp; English.
XX
CC   This protein comprises maize sucrose phosphate synthase (SPS), an
CC   enzyme that catalyses the formation of sucrose phosphate from
CC   fructose 6-phosphate and UDP-glucose in photosynthetically active
CC   plant cells.  It is considered to be a rate-limiting enzyme in the
CC   pathway providing sucrose to growing tissue.  The SPS amino acid
CC   sequence was deduced from cDNA clones (see AAT95847) derived from
CC   leaf cDNA.  The complete protein is believed to be a dimeric or
CC   tetrameric protein having a basic subunit of 110-130 kDa.  Peptides
CC   B11 (see AAW38269) and 4K (see AAW38270) isolated from SPS were used
CC   to design primers (see AAT95848-49) utilised in the isolation of
CC   SPS cDNA.  Transgenic plants that express SPS can be used to
CC   measure the effects on crop yield of an increased rate of sucrose
CC   translocation to growing tissues.
XX
SQ   Sequence   1068 AA;
```

Query Match 5.7%; Score 112.5; DB 18; Length 1068;
Best Local Similarity 20.9%; Pred. No. 0.1;
Matches 56; Conservative 56; Mismatches 87; Indels 69; Gaps 14;

```
Qy     125 GRVIPTFNVM---TDFCLHKIWVHENVDKYYVATDYVKEKLLEI-GTHPSNVKITGIPIR 180
           || :|   |:    ||    : |||::|        ||: :: : |  | ::
Db     427 grymprmvvippgmdf--snvvvhedid----gdgdvkddivglegaspksmp------- 473

Qy     181 PQFEESM-----PVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVV 235
           | : | |     |  |:    :    || :  :  | |  : ::||          : :
Db     474 piwaevmrfltnphkpmilalsrpdpkknittlvkafgecrplrel---------anltl 524

Qy     236 VCGKNTALKESLSALEAENGDKL-KVLGYVERID----------------ELFRITDCM 277
           : |     :: :  : | |   |  ||  ::: |                 |::|:   |
Db     525 imgn----rddiddmsagnasvlttvlklidkydlygsvafpkhhnqadvpeiyrlaakm 580
```

```
Qy    278 --------ITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEE-- 327
                  : :| |:|| || | |:|::  |     :  ||         ::|: |::
Db    581 kgvfinpalvepfgltlieaaahglpivatknggpvditna-----lnngllvdphdqna 635

Qy    328 ILESVTSLLADEDTLHRMKKN-IKDLHL 354
          | :::  |:||::     ::| ::::||
Db    636 iadallklvadknlwqecrrnglrnihl 663


RESULT 15
AAW09869
ID   AAW09869 standard; Protein; 1068 AA.
XX
AC   AAW09869;
XX
DT   25-JUL-1997  (first entry)
XX
DE   Sucrose phosphate synthase.
XX
KW   Sucrose phosphate synthase; sweetness; transgenic plant; tomato;
KW   fruit; tuber; sugar; maize; corn.
XX
OS   Zea mays hybrid 3184.
XX
PN   WO9715678-A2.
XX
PD   01-MAY-1997.
XX
PF   25-OCT-1996;   96WO-US17351.
XX
PR   27-OCT-1995;   95US-0549016.
XX
PA   (CALJ ) CALGENE INC.
XX
PI   Shewmaker CK;
XX
DR   WPI; 1997-259030/23.
DR   N-PSDB; AAT66229.
XX
PT   Modifying sweetness of plant parts by introducing transgene that
PT   encodes sucrose phosphate synthase - also altering ratio of soluble
PT   solids in sink tissue, especially for increasing sugar content in
PT   tomatoes
XX
PS   Example 5.6; Page 69-75; 112pp; English.
XX
CC   The amino acid sequence (AAW09869) of maize hybrid 3184 sucrose
CC   phosphate synthase (SPS) was deduced from cDNA clones (see also
CC   AAT66229) isolated from a leaf cDNA library.  SPS is considered a
CC   rate-limiting enzyme in the pathway providing sucrose to growing
CC   tissue.  Methods are provided for modifying the sweetness of plant
CC   sink tissue (partic. tomato fruit) in which SPS activity and/or
CC   invertase activity in plant tissues is manipulated.  Carbohydrate
CC   partitioning can be modified in plant tissues and/or parts, which
CC   in turn can be used to alter plant growth, soluble solid content
CC   and/or sweetness, and/or to alter the sensitivity of plant growth
CC   to temperature and/or to levels of carbon dioxide and oxygen.
XX
SQ   Sequence   1068 AA;


  Query Match             5.7%;  Score 112.5;  DB 18;  Length 1068;
  Best Local Similarity   20.9%;  Pred. No. 0.1;
  Matches   56;  Conservative   56;  Mismatches   87;  Indels   69;  Gaps   14;

Qy    125 GRVIPTFNVM---TDFCLHKIWVHENVDKYYVATDYVKEKLLEI-GTHPSNVKITGIPIR 180
          || :|   |:    ||    : |||::|        ||: :: : |  | ::
Db    427 grymprmvvippgmdf--snvvvhedid----gdgdvkddivglegaspksmp------- 473

Qy    181 PQFEESM-----PVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVV 235
          | : | |     |  |:    :    || :  :  | |  : ::||          : :
Db    474 piwaevmrfltnphkpmilalsrpdpkknittlvkafgecrplrel---------anltl 524
```

```
Qy    236 VCGKNTALKESLSALEAENGDKL-KVLGYVERID----------------ELFRITDCM 277
          : |     :: :  : | |   |  ||  ::: |                |::|:   |
Db    525 imgn----rddiddmsagnasvlttvlklidkydlygsvafpkhhnqadvpeiyrlaakm 580

Qy    278 --------ITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEE-- 327
                  : :| |:|| || | |:|::  |     :  ||         ::|: |::
Db    581 kgvfinpalvepfgltlieaaahglpivatknggpvditna-----lnngllvdphdqna 635

Qy    328 ILESVTSLLADEDTLHRMKKN-IKDLHL 354
          | :::  |:||::     ::| ::::||
Db    636 iadallklvadknlwqecrrnglrnihl 663


Search completed: June 29, 2001, 08:59:02
Job time: 175 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 29, 2001, 08:59:43 ; Search time 38.9 Seconds
(without alignments)
1299.242 Million cell updates/sec

Title: US-09-668-788-2
Perfect score: 1970
Sequence: 1 MNTNKRVLILTANYGNGHVQ.....EDILKESEMMTAKQKAKVLS 382

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	591.5	30.0	391	2	O86492	O86492 staphylococ
2	368	18.7	374	2	Q9KBH0	Q9kbh0 bacillus ha
3	353	17.9	411	2	Q9RVF3	Q9rvf3 deinococcus
4	326	16.5	522	10	Q9SM44	Q9sm44 spinacia ol
5	317	16.1	464	10	Q9SI93	Q9si93 arabidopsis
6	317	16.1	465	10	Q9FZL5	Q9fz15 arabidopsis
7	309	15.7	373	2	O34625	O34625 bacillus su
8	303.5	15.4	525	10	P93115	P93115 cucumis sat
9	302	15.3	530	10	Q9FZL4	Q9fzl4 glycine max
10	301.5	15.3	533	8	Q9MU68	Q9mu68 arabidopsis
11	301.5	15.3	533	10	O81770	O81770 arabidopsis
12	295	15.0	535	10	Q9FZL3	Q9fzl3 nicotiana t
13	293	14.9	468	10	O82730	O82730 arabidopsis
14	216	11.0	384	2	Q9EX00	Q9ex00 streptomyce
15	178	9.0	363	2	Q9K9T0	Q9k9t0 bacillus ha
16	163	8.3	383	2	O51410	O51410 borrelia bu
17	156.5	7.9	398	1	Q58652	Q58652 methanococc
18	138	7.0	378	2	Q9X0C4	Q9x0c4 thermotoga
19	136.5	6.9	458	10	O23406	O23406 arabidopsis
20	130.5	6.6	392	2	O34539	O34539 bacillus su
21	130.5	6.6	474	10	O04930	O04930 arabidopsis
22	128	6.5	406	2	Q9WZK3	Q9wzk3 thermotoga
23	126.5	6.4	374	2	Q9L9K6	Q919k6 pasteurella
24	125.5	6.4	372	2	Q9L4H3	Q914h3 brevibacter
25	123.5	6.3	367	2	Q9PF81	Q9pf81 xylella fas
26	122.5	6.2	342	2	Q9PNQ2	Q9pnq2 campylobact
27	122.5	6.2	1305	10	Q9FJ35	Q9fj35 arabidopsis
28	120.5	6.1	565	3	Q9HFH9	Q9hfh9 pichia etch
29	119	6.0	418	2	Q9RWP0	Q9rwp0 deinococcus
30	119	6.0	478	10	Q9S7R8	Q9s7r8 arabidopsis
31	117	5.9	400	1	O26432	O26432 methanobact
32	117	5.9	401	2	P95705	P95705 staphylococ
33	116.5	5.9	338	2	Q9ZDB8	Q9zdb8 rickettsia
34	116.5	5.9	361	2	Q9Z636	Q9z636 streptococc
35	116.5	5.9	366	2	O66840	O66840 aquifex aeo
36	116	5.9	363	1	Q9YCS0	Q9ycs0 aeropyrum p
37	115.5	5.9	571	1	Q58750	Q58750 methanococc
38	114.5	5.8	361	2	O07871	O07871 streptococc
39	114.5	5.8	404	2	P73369	P73369 synechocyst
40	114.5	5.8	468	10	Q9SBQ2	Q9sbq2 petunia hyb
41	112.5	5.7	479	10	Q9LSY6	Q9lsy6 arabidopsis
42	112.5	5.7	530	11	P97886	P97886 rattus norv
43	112.5	5.7	2890	2	O25806	O25806 helicobacte
44	112	5.7	769	10	O49978	O49978 actinidia c
45	111.5	5.7	391	2	P72382	P72382 staphylococ

ALIGNMENTS

RESULT 1
O86492
ID O86492 PRELIMINARY; PRT; 391 AA.
AC O86492;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE YPFP PROTEIN.
GN YPFP.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=98313013; PubMed=9650993;
RA Ludovice A.M., Wu S., de Lencastre H.;
RT "Molecular cloning and DNA sequencing of the Staphylococcus aureus
RT UDP-N-acetylmuramyl tripeptide synthetase (murE) gene, essential for
RT the optimal expression of methicillin resistance.";
RL Microb. Drug Res. 4:85-90(1998).
DR EMBL; Y14370; CAA74741.1; -.
SQ SEQUENCE 391 AA; 44703 MW; 40CDDC37A2627C6C CRC64;

Query Match 30.0%; Score 591.5; DB 2; Length 391;
Best Local Similarity 35.7%; Pred. No. 3.8e-30;
Matches 136; Conservative 77; Mismatches 131; Indels 37; Gaps 10;

```
Qy       4 NKRVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTV--SNLYQESNPIVSEVTQYLYLKS 61
           ||::||:| ::|||||:|| ::: |   :   |::|   :|: |::||:: : :  |: |
Db       5 NKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINS 64

Qy      62 FSIGKQFYRLFYYG----VDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVV 117
           |   :  |: |||     :|| : |     |: : || :  |: | :||:|: |||  |:
Db      65 FKYFRNMYKGFYYSRPDKLDKCFYK-----YYGL-NKLINLLIKE-KPDLILLTFPTPVM 117

Qy     118 PEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKITGI 177
                :   ||   ||||: ||| |:     :|||||   |:  :::|  || ||:|||
Db     118 SVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGI 177
```

```
Qy    178 PIRPQFEESMPVGPIYKKY-----NLSPNKKVLLIMAGAHGVLKNVKELCEN-LVKDDQV 231
          ||  :||      || :|      || |:|: :|: ||| || |    :  : | |
Db    178 PIDNKFET-----PINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSANA 232

Qy    232 QVVVVCGKNTALKESLSALEAENGDKLK------VLGYVERIDELFRITDCMITKPGGIT 285
          |||::|||:  || ||:|       | |      :||| : ::|    :  |||||||||
Db    233 QVVMICGKSKELKRSLTA-------KFKLTRMYLILGYTKHMNEWMASSQLMITKPGGIT 285

Qy    286 LTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDTLHRM 345
          :||  |  :|:|   | |||| ||| :||::|   : :  || :: | ||    : |  |
Db    286 ITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTNGNEQLTNM 345

Qy    346 KKNIKDLHLANSSEVILEDIL 366
            ::   :  ::: |  |:|
Db    346 ISTMEQDKIKYATQTICRDLL 366


RESULT   2
Q9KBH0
ID   Q9KBH0      PRELIMINARY;      PRT;   374 AA.
AC   Q9KBH0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BH1957 PROTEIN.
GN   BH1957.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RA   Takami H., Nakasone K., Takaki Y.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP001513; BAB05676.1; -.
SQ   SEQUENCE   374 AA;  42361 MW;  DA96A0EA33C5AC65 CRC64;


  Query Match            18.7%;  Score 368;  DB 2;  Length 374;
  Best Local Similarity  27.0%;  Pred. No. 6.1e-16;
  Matches  103;  Conservative   84;  Mismatches  159;  Indels   36;  Gaps   13;

Qy      8 LILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLKSFSIGKQ 67
          || :|: |||| | || |  :    |:|   :   |  | | : :     |:    :| :
Db      6 LIFSASIGNGHNQAAKALQVEFQNKGYQPEIIDTFYSLS-PALHKFMLTSYVNLLKVGPR 64

Qy     68 FYRLFYYGVDK---IYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEYRRRT 124
           ::  |:  :|        :|  :|    : |   |  ::   :::| | :     | ::
Db     65 IWQKIYFQAEKYPLFLFLDQFATFFV---ESLHATVKSNRCSFLVSTHPFVTAFLVRLKS 121

Qy    125 GRV--IPTFNVMTDFCLHKIWVHENVDKYYVA----TDYVKEKLLEIGTHPSNVK----- 173
           :   :| : |:||| ||  ::   :| |: :    ||: |          :||
Db    122 KKQLNLPLYTVITDFVLHPAYLRPEIDGYFTSDPNFTDFAK---------LNNVSDDRFF 172

Qy    174 ITGIPIRPQFEE-SMPVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENL-VKDDQV 231
           |||||  |  |    |   :     | | :|||:|  |  | | |  ::   |    : :
Db    173 PTGIPI-PNLESIDQPKWKVRNDLGLDPQRKVLIIAGGGIG-LTNYAQVIRALECLPEPI 230

Qy    232 QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGITLTEATA 291
          |:: : | |  :|| :| ::::  :|||: : ::     : :| :::| ||:|: |:
Db    231 QLLCMIGHNYQVKEKISRIKSKH--ELKVIEFTDKFLLYLKASDAILSKAGGLTMAESLV 288

Qy    292 IGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDTLHRMKKNIKD 351
             |:|:::|||| |: || |  | |||: |   :||  ::  :| :|     | :| :
Db    289 CETPIIIHQPVPGHEEHNAKFLIDAGAALRVKGSKEIPTTIKRVLYEEACFGPMIENARK 348

Qy    352 LHLANSSEVILEDIL---KESE 370
          |   |::  |:| :|   || :
Db    349 LKKPNAANEIVEQMLLLVKEQQ 370


RESULT   3
```

```
Q9RVF3
ID   Q9RVF3      PRELIMINARY;      PRT;   411 AA.
AC   Q9RVF3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CELL WALL SYNTHESIS PROTEIN, PUTATIVE.
GN   DR1076.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001958; AAF10649.1; -.
DR   TIGR; DR1076; -.
SQ   SEQUENCE   411 AA;  45017 MW;  D825DE52B8801437 CRC64;


  Query Match            17.9%;  Score 353;  DB 2;  Length 411;
  Best Local Similarity  26.2%;  Pred. No. 6.3e-15;
  Matches  101;  Conservative   77;  Mismatches  164;  Indels   44;  Gaps   11;

Qy      6 RVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYL--YLKSFS 63
          | | :: : | || |      :| |:  |    |  |  | : :     :||  : :||:
Db     29 RALFMSVSLGAGHDQA-----QQAVKQAFAERGVELLGAEHDSV-----EYLSTFERSFT 78

Qy     64 I---------GKQFYRLFYYGVDKIYNKRKFNIYFKM----GNKRLGELVDEHQPDIIIN 110
          :            || ||:  |:    : :||  :|    |     : : | :|:::||
Db     79 VDLYEFELRYAPWLYRGFYWLTDQ---DQPWNIISRMFTWLGMGAFKDELRLRPEVVIN 135

Qy    111 TF--PMIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIGTH 168
          :|  |  |    | :||:      ::||:  |  |     |   ||::  : ::|| |
Db    136 SFWAPAAVCDTLRAQTGQRFLNCLIVTDYRAHLHWARRETDLLMVASEETRRQMLERGVR 195

Qy    169 PSNVKITGIPIRPQFEESMPV------GPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELC 222
          |  |::|||| | | | :          :: : :| |   :||:  |  |      ::
Db    196 PEQVEVTGIPISPAFREVLAADRWALRAELFSEMSLRPGVPLLLLSGGGRGHYAAAADVL 255

Qy    223 ENLVK-DDQVQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKP 281
          |      |||:|   :     |::       |  :  ||:   : |  :| :: |
Db    256 TELGNLGRAVQVLVPASRQGEGTETI------GGATVHHLGFRRDLPRLLAASDLVVGKA 309

Qy    282 GGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDT 341
          ||:|: ||||:|||:::| |:||||: ||:| |  || :      ::   |   | |
Db    310 GGLTVAEATALGVPLVIYAPIPGQEEHNADFLERHGAGLWARARHDVRPLVLRAL-DPAE 368

Qy    342 LHRMKKNIKDLHLANSSEVILEDILK 367
            |:    : : : :::: :   ||:
Db    369 HARLSAGARAVGIPDAADRVAGAILR 394


RESULT   4
Q9SM44
ID   Q9SM44      PRELIMINARY;      PRT;   522 AA.
AC   Q9SM44;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MGDG SYNTHASE A PRECURSOR (EC 2.4.1.46).
GN   MGD A.
OS   Spinacia oleracea (Spinach).
```

```
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
OC   Caryophyllales; Chenopodiaceae; Spinacia.
OX   NCBI_TaxID=3562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99449603; PubMed=10518794;
RA   Miege C., Marechal E., Shimojima M., Awai K., Block M.A., Ohta H.,
RA   Takamiya K.I., Douce R., Joyard J.;
RT   "Biochemical and topological properties of type A MGDG synthase, a
RT   spinach chloroplast envelope enzyme catalyzing the synthesis of both
RT   prokaryotic and eukaryotic MGDG.";
RL   Eur. J. Biochem. 265:990-1001(1999).
DR   EMBL; AJ249607; CAB56218.1; -.
KW   Transit peptide; Transferase; Glycosyltransferase.
FT   TRANSIT       1     98
FT   CHAIN        99    522       MGDG SYNTHASE A.
SQ   SEQUENCE   522 AA;  57511 MW;  02E2B929732551A7 CRC64;


  Query Match             16.5%;  Score 326;  DB 10;  Length 522;
  Best Local Similarity   25.3%;  Pred. No. 4.6e-13;
  Matches   98;  Conservative   80;  Mismatches  162;  Indels   48;  Gaps   12;

Qy       5 KRVLILTANYGNGHVQVAKTLYEQCVRLGFQ-------HVTVSNLYQESN--PIVSEVTQ 55
           |||||| :: | ||   |     : ::  |          | |::|: |    |
Db     131 KRVLILMSDTGGGHRASA-----EAIKAAFNEEFGDDYQVFVTDLWSEHTPWPFNQLPRS 185

Qy      56 YLYLKSFSIGKQFYRLFYYGVD-KIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM 114
           | :|         ::: |||   :: ::  |         : :   : ::||||||:  |:
Db     186 YNFLVKHG---PLWKMMYYGTSPRVIHQSNFAATSVFIAREVARGLMKYQPDIIISVHPL 242

Qy     115 IVVPEYRRRTGR----VIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHP 169
           :     |   ||      |    |:||    |  | |: | : |  :: | ::  : |  |
Db     243 MQHVPLRILRGRGLLEKIVFTTVVTDLSTCHPTWFHKLVTRCYCPSNEVAKRATKAGLQP 302

Qy     170 SNVKITGIPIRPQFEESM-PVGPIYKKYNLSPNKKVLLIMAGAHGV--LKNVKELCENLV 226
           | :|: |:|:|| |  |: |   : |:}:  :   :|:| |  |:  ::       | :
Db     303 SQIKVYGLPVRPSFVRSVRPKNELRKELGMDEHLPAVLLMGGGEGMGPIEATARALGNAL 362

Qy     227 KDDQV-----QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKP 281
             |  :     |::|:||:|  |   ||::: :    ::| |:| :|:|     ||:|||
Db     363 YDANLGEPTGQLLVICGRNKKLAGKLSSIDWK--IPVQVKGFVTKIEECMGACDCIITKA 420

Qy     282 GGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDT 341
           |  |: ||   |:|:||   : |||  |  :  : |    :   :|| ::|:     :
Db     421 GPGTIAEAMIRGLPIILNDYIAGQEAGNVPYVIENGIGKYLKSPKEIAKTVSQWFGPK-- 478

Qy     342 LHRMKKNIKDLHLANSSEVILEDILKES 369
                     ||  ::: :: || :
Db     479 ------------ANELQIMSQNALKHA 493


RESULT   5
Q9SI93
ID   Q9SI93      PRELIMINARY;      PRT;   464 AA.
AC   Q9SI93;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PUTATIVE MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE.
GN   AT2G11810.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA   Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA   Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA   Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA   Salzberg S.L., Fraser C.M., Venter J.C.;
RT   "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL   Nature 402:761-768(1999).
DR   EMBL; AC007187; AAD28678.1; -.
DR   InterPro; IPR001296; -.
DR   Pfam; PF00534; Glycos_transf_1; 1.
SQ   SEQUENCE   464 AA;  52859 MW;  FEC2B424CFBBA136 CRC64;


  Query Match             16.1%;  Score 317;  DB 10;  Length 464;
  Best Local Similarity   25.6%;  Pred. No. 1.5e-12;
  Matches  101;  Conservative   90;  Mismatches  171;  Indels   32;  Gaps   16;

Qy       5 KRVLILTANYGNGHVQVAKTLYEQCVRLGF---QHVTVSNLYQESN--PIVSEVTQYLYL 59
           | |||| :: | ||   |: :    :: |    : : ::::|    |:     || ::
Db      72 KTVLILMSDTGGGHRASAEAI-RDAFKIEFGDDYRIIIKDVWKEYTGWPLNDMERQYKFM 130

Qy      60 KSFSIGKQFYRLFYYGVD-KIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM---- 114
              :|   : : ::|    |  :|   :        | :   : |::||||||:  |:
Db     131 VK-HVG--LWSVAFHGTSPKWIHKSYLSALAAYYAKEIEAGLMEYKPDIIISVHPLMQHI 187

Qy     115 -IVVPEYRRRTGRVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNV 172
            : | :::    :||    |:||    |: | |  | : |  :  | :: |  |   | :
Db     188 PLWVMKWQGLHKKVI-FVTVITDLNTCHRTWFHHGVSRCYCPSKEVAKRALVDGLDDSQI 246

Qy     173 KITGIPIRPQFEES-MPVGPIYKKYNLSPNKKVLLIMAGAHG---VLKNVKELCENLVKD 228
           :: |:|:|| |  : :    : |:  :  |   :|:| |  |   | |    | :.|
Db     247 RVFGLPVRPSFPRTILNKNELRKELEIDLNLPAVLLMGGGEGMGPVQKTALALGDSLYNS 306

Qy     229 DQV----QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGI 284
           :      |::|:||:|  |  :|:: | :    :|| |:  ::::     ||:||| |
Db     307 KESNPIGQLIVICGRNKVLASTLASHEWK--IPVKVRGFETQMEKWMGACDCIITKAGPG 364

Qy     285 TLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLA-DEDTLH 343
           |: ||   |:|:||   :||||| |  :  | || :     :|  : |    : ::: |
Db     365 TIAEALICGLPIILNDYIPGQEKGNVPYVVDNGAGVFTRSPKETAKIVADWFSNNKEELK 424

Qy     344 RMKKNIKDLHLANSSEVILEDILKESEMMTAKQK 377
           :| :|   | |:    |   ||:|:   :: :|:
Db     425 KMSENA--LKLSQPEAVF--DIVKDIHHLSQQQQ 454


RESULT   6
Q9FZL5
ID   Q9FZL5      PRELIMINARY;      PRT;   465 AA.
AC   Q9FZL5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MGDG SYNTHASE TYPE C.
GN   MGDC.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Awai K., Marechal E., Block M.A., Takamiya K., Joyard J., Ohta H.;
RT   "The Multigenic Family of MGDG synthases.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB047398; BAB12041.1; -.
SQ   SEQUENCE   465 AA;  52990 MW;  05E0157012E50A14 CRC64;


  Query Match             16.1%;  Score 317;  DB 10;  Length 465;
  Best Local Similarity   25.6%;  Pred. No. 1.5e-12;
  Matches  101;  Conservative   90;  Mismatches  171;  Indels   32;  Gaps   16;
```

```
Qy       5 KRVLILTANYGNGHVQVAKTLYEQCVRLGF---QHVTVSNLYQESN--PIVSEVTQYLYL 59
Db      73 KTVLILMSDTGGGHRASAEAI-RDAFKIEFGDDYRIIIKDVWKEYTGWPLNDMERQYKFM 131

Qy      60 KSFSIGKQFYRLFYYGVD-KIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM---- 114
Db     132 VK-HVG--LWSVAFHGTSPKWIHKSYLSALAAYYAKEIEAGLMEYKPDIIISVHPLMQHI 188

Qy     115 -IVVPEYRRRTGRVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNV 172
Db     189 PLWVMKWQGLHKKVI-FVTVITDLNTCHRTWFHHGVSRCYCPSKEVAKRALVDGLDDSQI 247

Qy     173 KITGIPIRPQFEES-MPVGPIYKKYNLSPNKKVLLIMAGAHG---VLKNVKELCENLVKD 228
Db     248 RVFGLPVRPSFPRTILNKNELRKELEIDLNLPAVLLMGGGEGMGPVQKTALALGDSLYNS 307

Qy     229 DQV----QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGI 284
Db     308 KESNPIGQLIVICGRNKVLASTLASHEWK--IPVKVRGFETQMEKWMGACDCIITKAGPG 365

Qy     285 TLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLA-DEDTLH 343
Db     366 TIAEALICGLPIILNDYIPGQEKGNVPYVVDNGAGVFTRSPKETAKIVADWFSNNKEELK 425

Qy     344 RMKKNIKDLHLANSSEVILEDILKESEMMTAKQK 377
Db     426 KMSENA--LKLSQPEAVF--DIVKDIHHLSQQQQ 455


RESULT   7
O34625
ID   O34625      PRELIMINARY;      PRT;   373 AA.
AC   O34625;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   YKON.
GN   YKON.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Devine K.M.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ002571; CAA05612.1; -.
DR   EMBL; Z99111; CAB13208.1; -.
DR   EMBL; Z99110; CAB13192.1; -.
SQ   SEQUENCE   373 AA;  43480 MW;  62FFCF2BDC598CDC CRC64;


  Query Match             15.7%;  Score 309;  DB 2;  Length 373;
  Best Local Similarity   24.6%;  Pred. No. 3.5e-12;
  Matches   99;  Conservative   89;  Mismatches  158;  Indels   56;  Gaps   16;

Qy       5 KRVLILT-ANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLKSFS 63
Db       2 KNILIFPFLSISTGHHHVADALQAELESQGLAAEKI-DIFSHSYRRLEKLSSVAYLKWI- 59

Qy      64 IGKQFYRLFYYGVDKI-------YNKRKFNIYFKMGNKRLGELVDEHQPDIIINT--FPM 114
Db      60 ---QYFPKTYSGIYRLLACGEFQHDKRYF-MYECVFTQQMRHILQEKQPDIAFCTHALPS 115

Qy     115 IVV----PEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIGTHPS 170
Db     116 YLLNRLKPEYPNLT-----VVNVYTDFFVNQLWGRKNIDYHFVPSTEVKKQLISEGIDQN 170

Qy     171 NVKITGIPIRPQFE----ESMPVGPIYKKYNLSPNKKVLLIMAGAHGV---LKNVKELCE 223
Db     171 NIYLTGIPVHQNFEMESADTLQHHPPY----------TIIITGGSMGVGGILKWVQELSP 220

Qy     224 NLVKDDQVQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGG 283
Db     221 G----GKILYKILCGRNEKLYSYVKSLHHPLIEAIPYLHSKAEMNRLYEQATGIMTKPGG 276

Qy     284 ITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRH----EEILESVTSLLADE 339
Db     277 VTISECLQKRLPVFIYHALPGQEEMNLNLLHER--KLVTDMRNWDMQKAEEYIAAFFQSN 334

Qy     340 DTLHRMKKNIKDLHLANSSEVILEDILKESEMMTAKQKAKVL 381
Db     335 EQMKEYKKHVNG-YLGEMSDRKIKDVLKR---IIWKQKNTLL 372


RESULT   8
P93115
ID   P93115      PRELIMINARY;      PRT;   525 AA.
AC   P93115;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE PRECURSOR (EC 2.4.1.46)
DE   (1,2-DIACYLGLYCEROL 3-BETA-GALACTOSYLTRANSFERASE).
OS   Cucumis sativus (Cucumber).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Cucurbitales; Cucurbitaceae; Cucumis.
OX   NCBI_TaxID=3659;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=AONAGAJIBAI;
RX   MEDLINE=97144442; PubMed=8990209;
RA   Shimojima M., Ohta H., Iwamatsu A., Masuda T., Shioi Y.,
RA   Takamiya K.-I.;
```

```
RT   "Cloning of the gene for monogalactosyldiacylglycerol synthase and its
RT   evolutionary origin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:333-337(1997).
CC   -!- FUNCTION: CATALYZES THE FORMATION OF MONOGALACTOSYLDIACYLGLYCEROL
CC       (MGDG) WHICH IS A MAJOR STRUCTURAL LIPID OF THE CHLOROPLAST.
CC   -!- CATALYTIC ACTIVITY: UDP-GALACTOSE + 1,2-DIACYLGLYCEROL = UDP + 3-
CC       BETA-D-GALACTOSYL-1,2- DIACYLGLYCEROL.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST.
CC   -!- SIMILARITY: TO E.COLI AND B.SUBTILIS UDP-N-ACETYLGLUCOSAMINE--N-
CC       ACETYLMURAMYL-(PENTAPEPTIDE) PYROPHOSPHORYL-UNDECAPRENOL N-
CC       ACETYLGLUCOSAMINE TRANSFERASE.
DR   EMBL; U62622; AAC49624.1; -.
DR   Mendel; 12656; Cucsa;1856;12656.
KW   Transferase; Glycosyltransferase; Transit peptide; Chloroplast.
FT   TRANSIT       1    103       CHLOROPLAST.
FT   CHAIN       104    525       MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE.
SQ   SEQUENCE   525 AA;  57862 MW;  74FE586082EC48BA CRC64;


  Query Match             15.4%;  Score 303.5;  DB 10;  Length 525;
  Best Local Similarity   25.7%;  Pred. No. 1.3e-11;
  Matches  100;  Conservative   77;  Mismatches  179;  Indels   33;  Gaps   13;

Qy       5 KRVLILTANYGNGH---VQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLKS 61
           |||||| :: | ||    :  |  : :      :| | :::|: :  |          |
Db     136 KRVLILMSDTGGGHRASAEAIKAAFNEEFGNNYQ-VFITDLWTDHTPWPFNQLPRSYNFL 194

Qy      62 FSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM-----IV 116
             |  :   :|     |: ::  |        : : : : :::|||||:  |:      |
Db     195 VKHGTLWKMTYYVTAPKVIHQSNFAATSTFIAREVAKGLMKYRPDIIISVHPLMQHVPIR 254

Qy     117 VPEYRRRTGRVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKIT 175
           :   :    ::: |  |:||     |  | |: | : |  :  | :: |  |  || :|:
Db     255 ILRSKGLLNKIVFT-TVVTDLSTCHPTWFHKLVTRCYCPSTEVAKRALTAGLQPSKLKVF 313

Qy     176 GIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGAHG---VLKNVKELCENLVKDD 229
           |:|:|| | :  |: |   : |:  :  |   :|:| |  |   :       | | : |  ::
Db     314 GLPVRPSFVK--PIRPKIELRKELGMDENLPAVLLMGGGEGMGPIEATAKALSKALYDEN 371

Qy     230 Q----VQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGGIT 285
             ||:|:|| |  |   | ::: :    ::| |:| :::|     ||:||| |  |
Db     372 HGEPIGQVLVICGHNKKLAGRLRSIDWK--VPVQVKGFVTKMEECMGACDCIITKAGPGT 429

Qy     286 LTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADE-DTLHR 344
           : ||   |:|:||   : |||  |  :  : |         :||   |       : | |
Db     430 IAEAMIRGLPIILNDYIAGQEAGNVPYVVENGCGKFSKSPKEIANIVAKWFGPKADELLI 489

Qy     345 MKKNIKDLHLANSSEVI-----LEDILKE 368
           | :|   | ||    |      | :::|:
Db     490 MSQNA--LRLARPDAVFKIVHDLHELVKQ 516


RESULT   9
Q9FZL4
ID   Q9FZL4      PRELIMINARY;      PRT;   530 AA.
AC   Q9FZL4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MGDG SYNTHASE TYPE A.
GN   GMMGD A.
OS   Glycine max (Soybean).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Glycine.
OX   NCBI_TaxID=3847;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Awai K., Takamiya K., Ohta H.;
RT   "cDNA cloning of MGDG synthase from tobacco and soybean.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB047475; BAB11979.1; -.
SQ   SEQUENCE   530 AA;  57839 MW;  D33C37FD53E90218 CRC64;


  Query Match             15.3%;  Score 302;  DB 10;  Length 530;
  Best Local Similarity   27.0%;  Pred. No. 1.6e-11;
  Matches  104;  Conservative   79;  Mismatches  168;  Indels   34;  Gaps   17;

Qy       5 KRVLILTANYGNGH---VQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLKS 61
           |:|||| :: | ||    :  |  : |      :| | |::|: :  |          | |
Db     139 KKVLILMSDTGGGHRASAEAIKAAFYQEFGDDYQ-VFVTDLWADHTPWPFNQLPRSY--S 195

Qy      62 FSIGK-QFYRLFYYG-VDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMI--VV 117
           | :      ::: |||   :: ::  |        : : : : ::|||||:  |::  |
Db     196 FLVKHGPLWKMTYYGTAPRVVHQSNFAATGTFIAREVAKGLMKYQPDIIISVHPLMQHVP 255

Qy     118 PEYRRRTG--RVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKI 174
           |  |  : |    |:||     |  | |: | : |  |  | :: |: |   | :||
Db     256 LRILRSKGLLKKIVFTTVITDLSTCHPTWFHKLVTRCYCPTTDVAQRALKAGLQQSQIKI 315

Qy     175 TGIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGAHG---VLKNVKELCENLVKD 228
           |:|:|| | :  || |   : ::  :  :   :|:| |  |   :     : | ::|  |
Db     316 FGLPVRPSFVK--PVQPKDELRRELGMDEDLPAVLLMGGGEGMGPIEATARALGDSLY-D 372

Qy     229 DQV-----QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPGG 283
           : :     |::|:||:|  |   ||::  :    ::| |:| :::|     ||:||| |
Db     373 ENIGAPVGQILVICGRNKKLANKLSSINWK--VPVQVKGFVTKMEECMGACDCIITKAGP 430

Qy     284 ITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADED-TL 342
           |: ||   |:|:||   : |||  |  :  : |         ::| : |       :   |
Db     431 GTIAEAQIRGLPIILNDYIAGQEAGNVPYVVENGCGKFSKSPKDIAKIVAEWFGPKAYEL 490

Qy     343 HRMKKNIKDLHLANSSEV--ILEDI 365
           :| :|   | ||    |  |: |:
Db     491 QQMSQNA--LRLARPDAVFKIVHDL 513


RESULT  10
Q9MU68
ID   Q9MU68      PRELIMINARY;      PRT;   533 AA.
AC   Q9MU68;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE.
GN   MGD1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Jarvis P., Doermann P., Peto C.A., Lutes J., Benning C., Chory J.;
RT   "Galactolipid-Deficiency and Abnormal Chloroplast Development in the
RT   Arabidopsis MGD Synthase 1 Mutant.";
RL   Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2000).
DR   EMBL; AF241797; AAF65066.1; -.
KW   Chloroplast.
SQ   SEQUENCE   533 AA;  58537 MW;  E581E67317CB9CC8 CRC64;


  Query Match             15.3%;  Score 301.5;  DB 8;  Length 533;
  Best Local Similarity   24.8%;  Pred. No. 1.7e-11;
  Matches  101;  Conservative   86;  Mismatches  172;  Indels   49;  Gaps   17;

Qy       5 KRVLILTANYGNGHVQVAKTLYEQCVRLGFQ-------HVTVSNLYQESN--PIVSEVTQ 55
           |:|||| :: | ||   |      : :|  |          | :::|: :    |
Db     142 KKVLILMSDTGGGHRASA-----EAIRAAFNQEFGDEYQVFITDLWTDHTPWPFNQLPRS 196

Qy      56 YLYLKSFSIGKQFYRLFYYGVD-KIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM 114
           | :|         ::: |||   :| ::  |     .   : : : : ::||||||:  |:
```

```
Db    197 YNFLVKHG---TLWKMTYYGTSPRIVHQSNFAATSTFIAREIAQGLMKYQPDIIISVHPL 253

Qy    115 I--VVPEYRRRTG--RVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHP 169
          :  |      |  |  : |    |:||    |  | |: | : |  :  | ::  : |
Db    254 MQHVPLRVLRSKGLLKKIVFTTVITDLSTCHPTWFHKLVTRCYCPSTEVAKRAQKAGLET 313

Qy    170 SNVKITGIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGAHGV--LKNVKELCEN 224
          | :|: |:|:|| | :  || |   : ::  :  |   :|:| |  |:  ::       :
Db    314 SQIKVYGLPVRPSFVK--PVRPKVELRRELGMDENLPAVLLMGGGEGMGPIEATARALAD 371

Qy    225 LVKDDQV-----QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMIT 279
            : |  :     ||:::||:|  |:  ||:|: :    ::| |:: :::|     ||:||
Db    372 ALYDKNLGEAVGQVLIICGRNKKLQSKLSSLDWK--IPVQVKGFITKMEECMGACDCIIT 429

Qy    280 KPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLAD- 338
          | |  |: ||   |:|:||   : |||  |  :  : |      :  :  : :: ::||
Db    430 KAGPGTIAEAMIRGLPIILNGYIAGQEAGNVPYVVENGC----GKFSKSPKEISKIVADW 485

Qy    339 ----EDTLHRMKKNIKDLHLANSSEVILEDILKESEMMTAKQKAKVLS 382
                | | :|   | ||   | : : :    |::  |     ||
Db    486 FGPASKELEIMSQNA--LRLA-KPEAVFKIVHDMHELVRKKNSLPQLS 530


RESULT 11
O81770
ID   O81770      PRELIMINARY;      PRT;   533 AA.
AC   O81770;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE - LIKE PROTEIN
DE   (MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE-LIKE PROTEIN) (MGDG SYNTHASE
DE   TYPE A).
GN   F28M20.30 OR AT4G31780 OR MGDA.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Hoheisel J.,
RA   Mewes H.W., Mayer K.F.X., Schueller C., Bevan M.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Mewes H.W.,
RA   Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Terryn N., Ardiles W., Buysshaert C., Dasseville R., De Clerck R.,
RA   De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villaroel R.,
RA   Gielen J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Awai K., Marechal E., Block M.A., Takamiya K., Joyard J., Ohta H.;
RT   "The Multigenic Family of MGDG synthases.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031004; CAA19745.1; -.
DR   EMBL; AL161579; CAB79896.1; -.
DR   EMBL; AB047399; BAB12042.1; -.
DR   Mendel; 32401; Arath;1856;32401.
SQ   SEQUENCE   533 AA;  58537 MW;  E581E67317CB9CC8 CRC64;


  Query Match            15.3%;  Score 301.5;  DB 10;  Length 533;
  Best Local Similarity  24.8%;  Pred. No. 1.7e-11;
  Matches  101;  Conservative   86;  Mismatches  172;  Indels   49;  Gaps   17;

Qy      5 KRVLILTANYGNGHVQVAKTLYEQCVRLGFQ-------HVTVSNLYQESN--PIVSEVTQ 55
          |:|||| :: | ||   |     : :|  |          | :::|: :    |
Db    142 KKVLILMSDTGGGHRASA-----EAIRAAFNQEFGDEYQVFITDLWTDHTPWPFNQLPRS 196

Qy     56 YLYLKSFSIGKQFYRLFYYGVD-KIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM 114
          | :|         ::: |||   :| ::  |         : : : : ::||||||:  |:
Db    197 YNFLVKHG---TLWKMTYYGTSPRIVHQSNFAATSTFIAREIAQGLMKYQPDIIISVHPL 253

Qy    115 I--VVPEYRRRTG--RVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHP 169
          :  |      |  |  : |    |:||    |  | |: | : |  :  | ::  : |
Db    254 MQHVPLRVLRSKGLLKKIVFTTVITDLSTCHPTWFHKLVTRCYCPSTEVAKRAQKAGLET 313

Qy    170 SNVKITGIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGAHGV--LKNVKELCEN 224
          | :|: |:|:|| | :  || |   : ::  :  |   :|:| |  |:  ::       :
Db    314 SQIKVYGLPVRPSFVK--PVRPKVELRRELGMDENLPAVLLMGGGEGMGPIEATARALAD 371

Qy    225 LVKDDQV-----QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMIT 279
            : |  :     ||:::||:|  |:  ||:|: :    ::| |:: :::|     ||:||
Db    372 ALYDKNLGEAVGQVLIICGRNKKLQSKLSSLDWK--IPVQVKGFITKMEECMGACDCIIT 429

Qy    280 KPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLAD- 338
          | |  |: ||   |:|:||   : |||  |  :  : |      :  :  : :: ::||
Db    430 KAGPGTIAEAMIRGLPIILNGYIAGQEAGNVPYVVENGC----GKFSKSPKEISKIVADW 485

Qy    339 ----EDTLHRMKKNIKDLHLANSSEVILEDILKESEMMTAKQKAKVLS 382
                | | :|   | ||   | : : :    |::  |     ||
Db    486 FGPASKELEIMSQNA--LRLA-KPEAVFKIVHDMHELVRKKNSLPQLS 530


RESULT 12
Q9FZL3
ID   Q9FZL3      PRELIMINARY;      PRT;   535 AA.
AC   Q9FZL3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MGDG SYNTHASE TYPE A.
GN   NTMGD A.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC   Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Awai K., Takamiya K., Ohta H.;
RT   "cDNA cloning of MGDG synthase from tobacco and soybean.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB047476; BAB11980.1; -.
SQ   SEQUENCE   535 AA;  59589 MW;  B1B2067E86EDE477 CRC64;


  Query Match            15.0%;  Score 295;  DB 10;  Length 535;
  Best Local Similarity  26.2%;  Pred. No. 4.5e-11;
  Matches  102;  Conservative   76;  Mismatches  169;  Indels   42;  Gaps   15;

Qy      5 KRVLILTANYGNGHVQVAKTLYEQCVRLGFQ-------HVTVSNLYQESNPIVSEVTQYL 57
          |:|||| :: | ||   |     : :|  |          | :::|: |  |
Db    145 KKVLILMSDTGGGHRASA-----EAIRSAFNEEFGDKYQVFITDLWTEHTPWPFNQLPRS 199

Qy     58 YLKSFSIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMI-- 115
          |      | :   :|    :: ::  |         : : : : ::|||||:  |::
Db    200 YNFLVKHGSLWRMTYYATAPRLVHQTNFAATSTFIAREVAKGLMKYQPDIIISVHPLMQH 259

Qy    116 VVPEYRRRTG---RVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSN 171
          |      | |   ::| |  |:||    |  | |: | : |  :: | ::  |  |
Db    260 VPLRILRSKGLLKKIIFT-TVITDLSTCHPTWFHKLVTRCYCPSEEVAKRALRAGLKPYQ 318

Qy    172 VKITGIPIRPQFEESMPVGP---IYKKYNLSPNKKVLLIMAGAHG---VLKNVKELCENL 225
          :|: |:|:|| | :  || |   : |:  :  :   :|:| |  |   :    : | : |
Db    319 LKVYGLPVRPSFVK--PVPPKVELRKELGMEEHLPAVLLMGGGEGMGPIEATARALGDAL 376
```

```
Qy    226 VKDDQV------QVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMIT 279
          |::      ||:|:||:|  |   |:::: :    ::| |:| :::|     ||:||
Db    377 Y--DEIHGEPIGQVLVICGRNKKLFNRLTSVQWK--IPVQVKGFVTKMEECMGACDCIIT 432

Qy    280 KPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLAD- 338
          | |  |: ||   |:|:||   : |||  |  :  : |        ::|   |
Db    433 KAGPGTIAEAVIRGLPIILNDYIAGQEAGNVPYVIENGCGKFSKSPKKIANIVAQWFGPR 492

Qy    339 EDTLHRMKKNIKDLHLANSSEV--ILEDI 365
          :| |  | :|   | ||    |  |: |:
Db    493 QDELRIMSQNA--LRLARPDAVFKIVHDM 519


RESULT 13
O82730
ID   O82730      PRELIMINARY;      PRT;   468 AA.
AC   O82730;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE (EC 2.4.1.46).
GN   MGD.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Awai K., Shimojima M., Masuda T., Takamiya K.I., Ohta H.;
RT   "cDNA cloning of an Arabidopsis monogalactosyldiacylglycerol synthase
RT   revealed that the enzyme belongs to an uridine diphosphate-utilizing
RT   glycosyltransferase superfamily.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ000331; CAA04005.1; -.
DR   Mendel; 33300; Arath;1856;33300.
DR   InterPro; IPR000566; -.
DR   PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   468 AA;  52726 MW;  58F56C3C1C383C24 CRC64;


  Query Match             14.9%;  Score 293;  DB 10;  Length 468;
  Best Local Similarity   25.1%;  Pred. No. 5e-11;
  Matches   98;  Conservative   86;  Mismatches  172;  Indels   34;  Gaps   14;

Qy      5 KRVLILTANYGNGHVQVAKTLYEQC-VRLGFQH-VTVSNLYQESN--PIVSEVTQYLYLK 60
          | |||| :: | ||   |: : :   :  | :: | | ::::|    |:       | ::
Db     69 KNVLILMSDTGGGHRASAEAIRDAFKIEFGDKYRVIVKDVWKEYTGWPLNDMERSYKFMV 128

Qy     61 SFSIGKQFYRLFYYG-----VDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPM- 114
             | ::: ::     :   |      | |      |     |::|:|||:  |:
Db    129 KH---VQLWKVAFHSTSPKWIHSCYLAAIAAYYAKEVEAGL----MEYKPEIIISVHPLM 181

Qy    115 ----IVVPEYRRRTGRVIPTFNVMTDF-CLHKIWVHENVDKYYVATDYVKEKLLEIGTHP 169
              : | :::    ||:    |:||    |  | |  |:: |  :  | :: |  |
Db    182 QHIPLWVLKWQELQKRVL-FVTVITDLNTCHPTWFHPGVNRCYCPSQEVAKRALFDGLDE 240

Qy    170 SNVKITGIPIRPQFEESMPV-GPIYKKYNLSPNKKVLLIMAGAHGV--LKNVKELCENLV 226
          | |:: |:|:|| |  :: |   : |:  :  : : :|:| |  |:  :|   :  |  :
Db    241 SQVRVFGLPVRPSFARAVLVKDDLRKELEMDQDLRAVLLMGGGEGMGPVKETAKALEEFL 300

Qy    227 KDDQ-----VQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKP 281
            | :      |:||:||:|  |  :| |:: :    :|| |:  :::::     ||:|||
Db    301 YDKENRKPIGQMVVICGRNKKLASALEAIDWK--IPVKVRGFETQMEKWMGACDCIITKA 358

Qy    282 GGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADE-D 340
          |  |: |:    :|:||   :|||||  |  :  : || :     :|    |    : : |
Db    359 GPGTIAESLIRSLPIILNDYIPGQEKGNVPYVVENGAGVFTRSPKETARIVGEWFSTKTD 418

Qy    341 TLHRMKKNIKDLHLANSSEVILEDILKESE 370
```

```
              | :   | : |    :   |::|| : ||
Db    419 ELEQTSDNARKLAQPEAVFDIVKDIDELSE 448


RESULT 14
Q9EX00
ID   Q9EX00      PRELIMINARY;      PRT;   384 AA.
AC   Q9EX00;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PUTATIVE SECRETED PROTEIN.
GN   2SCG38.19C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Saunders D.C., Harris D.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL445503; CAC13078.1; -.
SQ   SEQUENCE   384 AA;  40689 MW;  E64E01E05B1DDE40 CRC64;


  Query Match             11.0%;  Score 216;  DB 2;  Length 384;
  Best Local Similarity   22.8%;  Pred. No. 3e-06;
  Matches   81;  Conservative   61;  Mismatches  167;  Indels   46;  Gaps   10;

Qy      5 KRVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTVSNLYQESNPIVSEVTQYLYLKSFSI 64
          :|:|:::|: | ||  ||  |  :    |    ||                 | |  : :
Db     15 RRLLVISASMGAGHDTVAAELVRRARERGDTAQTV--------------DVLALLPYGL 59

Qy     65 GKQFYRLFYYGVDKIY---NKRKFNIYFKMGNKR------LGELVDEHQPDIIINTFPMI 115
          |    | || |  : :        : :: : |  |      |  |  :   ::   |  :
Db     60 G-AVLRCFYRGSVRHFPWAYAALYRLFLRPGAGRRPSGTPLARLAGDRLRELAARTGADV 118

Qy    116 VVPEYR---RRTGRV-------IPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEI 165
          |||  :    : || :       :|:  :: || ||: |:|   |     |:    :
Db    119 VVPVFHLGAQLTGHLRDRGLLPVPSVVLVIDFELHRQWLHPGNDHCLCLTEEAAREAR-- 176

Qy    166 GTHPSNVKITGIPIRPQFEESMPVGPIYKKY---NLSPNKKVLLIMAGAHGVLKNVKELC 222
          |   :  :  |  : |:|      |    :     | | :  ::: ||| ||  :: :
Db    177 GNTGTPAETCGPVVAPEFSAGRVPGAAQWRETFDRLGPGRPAVVLSAGAWGVGSHL-DAT 235

Qy    223 ENLVKDDQVQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITKPG 282
            |: ||   ||:|| |  |: :||       :  ||:|  :  |      :|
Db    236 VRLLVDHGYLPVVLCGDNQRLRRTLSGTPG-----VLALGWVTDMPGLLHAARALIDNAA 290

Qy    283 GITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLA 337
          | |  :| | |:||: ::|:||   :         | : |       :||::  | |
Db    291 GQTAVQALAAGLPVVGHRPIPGHGADGVRRMAALGVSEVAEDRTALLEALARLTA 345


RESULT 15
Q9K9T0
ID   Q9K9T0      PRELIMINARY;      PRT;   363 AA.
AC   Q9K9T0;
```

```
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE-N-ACETYLMURAMYL-(PENTAPEPTIDE)PYROPHOS.
GN   MURG.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RA   Takami H., Nakasone K., Takaki Y.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP001515; BAB06284.1; -.
SQ   SEQUENCE   363 AA;  39715 MW;  3B1DD063E495EE1B CRC64;

  Query Match             9.0%;  Score 178;  DB 2;  Length 363;
  Best Local Similarity   23.1%;  Pred. No. 0.00074;
  Matches   72;  Conservative   62;  Mismatches  124;  Indels   54;  Gaps   14;

Qy      87 IYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHE 146
           : |  | ||   |::| :||::| |   :  |         |||             :||
Db      75 VRFLRGTKRAKALLNEFKPDVVIGTGGYVCGPVVYAAAKLKIPTV-----------IHE 122

Qy     147 N------VDKY---YVATDYVKEKLLEIGTHPSNVKITGIPIRPQF-----EESMPVGPI 192
                  :|:   ||    :  |  |     : |  || |   :       || :
Db     123 QNSVPGLTNKFLSRYVDRIAICFKEAEAFFPKNKVVFTGNPRASEVMSGNREEGL----- 177

Qy     193 YKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQV---QVVVVCG--KNTALKESL 247
           :   : |||| :||: |: |  : : :   ::: | :    | | | |      ::| :
Db     178 -RSLGIKPNKKTVLIVGGSRGA-RPINDAFMSILSDVKAKPYQFVYVTGTVHYERVQEQM 235

Qy     248 SAL-EAENGDKLKVLGYVERIDELFRITDCMITKPGGITLTEATAIGVPVILYKPVP--- 303
           :: : ||   : |  ::  : ::    | :: : |  || | ||:|:| ||  | |
Db     236 KSIGQPEN---VIVQPFIHNMPDVLSAVDLIVARAGATTLAEITALGLPSILI-PSPYVT 291

Qy     304 -GQEKENANFFEDRGAAIVVNRHE----EILESVTSLLADEDTLHRMKKNIKDLHLANSS 358
             :::||     : |||:    |     :|| :  ::     |  ||:  | | :  ::
Db     292 NNHQEKNAAALSKKDAAILRKESELTGDRLLEDIDDIMVTPGRLDAMKQAAKALGVPTAA 351

Qy     359 E---VILEDILK 367
           |   :::::: |
Db     352 EKLHMLVKEVAK 363


Search completed: June 29, 2001, 09:05:02
Job time: 319 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 29, 2001, 08:56:42 ; Search time 24.11 Seconds
(without alignments)
319.173 Million cell updates/sec

Title: US-09-668-788-2
Perfect score: 1970
Sequence: 1 MNTNKRVLILTANYGNGHVQ.....EDILKESEMMTAKQKAKVLS 382

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	117.5	6.0	380	2	US-08-846-762-86	Sequence 86, Appl
2	112.5	5.7	1068	2	US-08-429-054A-11	Sequence 11, Appl
3	112.5	5.7	1068	2	US-08-718-777-7	Sequence 7, Appli
4	112.5	5.7	1068	4	US-09-051-341-7	Sequence 7, Appli
5	110	5.6	529	5	PCT-US92-00282-7	Sequence 7, Appli
6	108	5.5	892	1	US-07-977-434-12	Sequence 12, Appl
7	108	5.5	892	1	US-08-458-819-12	Sequence 12, Appl
8	108	5.5	892	5	PCT-US91-07035-12	Sequence 12, Appl
9	105.5	5.4	1220	2	US-08-843-530B-36	Sequence 36, Appl
10	102.5	5.2	416	4	US-09-007-476-2	Sequence 2, Appli
11	101	5.1	403	2	US-08-533-669A-10	Sequence 10, Appl
12	101	5.1	403	2	US-08-607-509-2	Sequence 2, Appli
13	101	5.1	403	2	US-08-607-509-4	Sequence 4, Appli
14	101	5.1	403	2	US-08-454-036-2	Sequence 2, Appli
15	101	5.1	403	2	US-08-634-642-2	Sequence 2, Appli
16	101	5.1	403	2	US-08-634-642-4	Sequence 4, Appli
17	101	5.1	403	3	US-08-989-370-2	Sequence 2, Appli
18	101	5.1	403	3	US-08-989-370-4	Sequence 4, Appli
19	101	5.1	403	5	PCT-US95-05064-2	Sequence 2, Appli
20	99	5.0	376	2	US-08-846-762-85	Sequence 85, Appl
21	99	5.0	1045	2	US-08-553-436A-6	Sequence 6, Appli
22	98	5.0	531	5	PCT-US92-00282-5	Sequence 5, Appli
23	98	5.0	1057	4	US-08-853-948B-2	Sequence 2, Appli
24	97.5	4.9	578	1	US-08-766-014-4	Sequence 4, Appli
25	97.5	4.9	608	1	US-08-766-014-3	Sequence 3, Appli
26	96.5	4.9	846	1	US-08-356-354-2	Sequence 2, Appli
27	96.5	4.9	846	2	US-08-778-656-2	Sequence 2, Appli
28	96.5	4.9	872	1	US-08-766-014-2	Sequence 2, Appli
29	95.5	4.8	1054	1	US-08-356-354-4	Sequence 4, Appli
30	95.5	4.8	1054	2	US-08-778-656-4	Sequence 4, Appli
31	94	4.8	533	5	PCT-US92-00282-3	Sequence 3, Appli
32	93.5	4.7	2391	2	US-08-446-855A-2	Sequence 2, Appli
33	93.5	4.7	2391	4	US-09-150-741-2	Sequence 2, Appli
34	92.5	4.7	241	6	5213972-7	Patent No. 5213972
35	92.5	4.7	908	1	US-08-356-354-6	Sequence 6, Appli
36	92.5	4.7	908	2	US-08-778-656-6	Sequence 6, Appli
37	92	4.7	531	5	PCT-US92-00282-6	Sequence 6, Appli
38	90.5	4.6	507	1	US-08-097-829-6	Sequence 6, Appli
39	90.5	4.6	507	1	US-08-577-403-6	Sequence 6, Appli
40	90.5	4.6	751	2	US-08-843-530B-32	Sequence 32, Appl
41	90.5	4.6	1276	1	US-08-717-515-8	Sequence 8, Appli
42	90	4.6	876	1	US-08-717-515-4	Sequence 4, Appli
43	89	4.5	586	2	US-08-630-822A-70	Sequence 70, Appl
44	89	4.5	586	2	US-09-005-069-70	Sequence 70, Appl
45	88	4.5	521	1	US-08-090-523-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-08-846-762-86
; Sequence 86, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: de Kievit, Teresa
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
; TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa
; FILE REFERENCE: 6580-089
; CURRENT APPLICATION NUMBER: US/08/846,762A
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 86
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-08-846-762-86

Query Match 6.0%; Score 117.5; DB 2; Length 380;
Best Local Similarity 20.4%; Pred. No. 0.00072;
Matches 72; Conservative 69; Mismatches 169; Indels 43; Gaps 12;

```
Qy      48 PIVSEVTQYLYLKSF-SIGKQFYRLFYYGVDKIYNKRKFNIYFKMGNK----------R 95
           |:| |: :|  : |: ::  |  ::    :|  : |  |::      :          |
Db      21 PLVLELKKYPEIDSYVTVTAQHRQMLDQVLDAFHIKPDFDLNIMKERQTLAEITSNALVR 80

Qy      96 LGELVDEHQPDIII------NTFPMIVVPEYRR-RTGRV---IPTFNVMTDFC--LHKIW 143
           | ||  : :|||::      ||   :   | :  | |   : | | : |   |::
Db      81 LDELFKDIKPDIVLVHGDTTTTFAGSLAAFYHQIAVGHVEAGLRTGNKYSPFPEELNRQM 140

Qy     144 VHENVDKYYVATDYVKEKLLEIGTHPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKK 203
              ||::  |   |: ||:      :: :||         ::  |  :   :     |
Db     141 TGAIADLHFAPTGQAKDNLLKENKKADSIFVTGNTAIDALNTTVRDGYSHPVLDQVGEDK 200

Qy     204 VLLIMAGAHGVLKNVKELCENLVKD--------DQVQVVVVCGKNTALKESLSALEAENG 255
           ::|: |      :|: |  ||: |          : ||||     |  ::|: :     :
Db     201 MILLTAHRR---ENLGEPMENMFKAIRRIVGEFEDVQVVYPVHLNPVVREA-AHKHFGDS 256

Qy     256 DKLKVLGYVERID--ELFRITDCMITKPGGITLTEATAIGVPVILYKPVPGQEKENANFF 313
           |:: ::  :| ||      :  ::|  ||:  || ::| ||:: :      |
Db     257 DRVHLIEPLEVIDFHNFAAKSHFILTDSGGVQ-EEAPSLGKPVLVLR----DTTERPEGV 311

Qy     314 EDRGAAIVVNRHEEILESVTSLLADEDTLHRMKKNIKDLHLANSSEVILEDIL 366
           |     :     | | :    || | |   :| :         :|  |:|::|
Db     312 EAGTLKLAGTDEENIYQLAKQLLTDPDEYKKMSQASNPYGDGEASRRIVEELL 364
```

```
RESULT   2
US-08-429-054A-11
; Sequence 11, Application US/08429054A
; Patent No. 5917126
;  GENERAL INFORMATION:
;    APPLICANT:  VAN ASSCHE, CHARLES; LANDO, DANIELLE; BRUNEAU,
;    APPLICANT:  JEAN; VOELKER, TONI; GERVAIS, MONICA
;    TITLE OF INVENTION:  SUCROSE PHOSPHATE SYNTHETASE (SPS),
;    TITLE OF INVENTION:  A PREPARATION METHOD AND cDNA THEREFOR, AND USE OF THE
;    TITLE OF INVENTION:  cDNA FOR MODIFYING SPS EXPRESSION IN PLANT CELLS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BIERMAN AND MUSERLIAN
;      STREET:  600 THIRD AVENUE
;      CITY:  NEW YORK
;      STATE:  NEW YORK
;      COUNTRY:  USA
;      ZIP:  10016
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FLOPPY DISK
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/429,054A
;      FILING DATE:  26-APR-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US  842,337
;      FILING DATE:  20-March-1992
;      APPLICATION NUMBER:  PCT/FR 91/00593
;      FILING DATE:  18-July-1991
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  French 90402094.9
;      FILING DATE:  20-July-1990
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Charles A. Muserlian
;      REGISTRATION NUMBER:  19,683
;      REFERENCE/DOCKET NUMBER:  146.1137
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 661-8000
;      TELEFAX:  (212) 661-8002
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1068
;      TYPE:  Amino acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Unknown
;    MOLECULE TYPE:  Peptide
US-08-429-054A-11
```

```
  Query Match             5.7%;  Score 112.5;  DB 2;  Length 1068;
  Best Local Similarity   20.9%;  Pred. No. 0.012;
  Matches   56;  Conservative   56;  Mismatches   87;  Indels   69;  Gaps   14;

Qy     125 GRVIPTFNVM---TDFCLHKIWVHENVDKYYVATDYVKEKLLEI-GTHPSNVKITGIPIR 180
           || :|   |:    ||    : |||::|       ||: :: : |  | ::
Db     427 GRYMPRMVVIPPGMDF--SNVVVHEDID----GDGDVKDDIVGLEGASPKSMP------- 473

Qy     181 PQFEESM-----PVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVV 235
           | : | |     |  |:    :    || :  :  | |  : ::||          : :
Db     474 PIWAEVMRFLTNPHKPMILALSRPDPKKNITTLVKAFGECRPLREL---------ANLTL 524

Qy     236 VCGKNTALKESLSALEAENGDKL-KVLGYVERID----------------ELFRITDCM 277
           : |     :: :  : | |   |  ||  ::: |                |::|:   |
Db     525 IMGN----RDDIDDMSAGNASVLTTVLKLIDKYDLYGSVAFPKHHNQADVPEIYRLAAKM 580

Qy     278 --------ITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEE-- 327
```

```
                : :| |:|| || | |:|::  |     :  ||        ::|: |::
Db     581 KGVFINPALVEPFGLTLIEAAAHGLPIVATKNGGPVDITNA-----LNNGLLVDPHDQNA 635

Qy     328 ILESVTSLLADEDTLHRMKKN-IKDLHL 354
           | :::  |:||::     ::| ::::||
Db     636 IADALLKLVADKNLWQECRRNGLRNIHL 663
```

```
RESULT   3
US-08-718-777-7
; Sequence 7, Application US/08718777
; Patent No. 5981852
;  GENERAL INFORMATION:
;    APPLICANT:  Van Assche, C.
;    APPLICANT:  Lando, D.
;    APPLICANT:  Bruneau, J. M.
;    APPLICANT:  Voelker, T.
;    APPLICANT:  Gervais, M.
;    TITLE OF INVENTION:  MODIFICATION OF SUCROSE
;    TITLE OF INVENTION:  PHOSPHATE
;    TITLE OF INVENTION:  SYNTHASE IN PLANTS
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Law Offices of Barbara Rae-Venter
;      STREET:  260 Sheridan Avenue, Suite 440
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/718,777
;      FILING DATE:  NOT YET ASSIGNED
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/175,471
;      FILING DATE:  27-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barbara Rae-Venter
;      REGISTRATION NUMBER:  32,750
;      REFERENCE/DOCKET NUMBER:  CGNE.072.02US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)328-4400
;      TELEFAX:  (415)328-4477
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1068 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-718-777-7
```

```
  Query Match             5.7%;  Score 112.5;  DB 2;  Length 1068;
  Best Local Similarity   20.9%;  Pred. No. 0.012;
  Matches   56;  Conservative   56;  Mismatches   87;  Indels   69;  Gaps   14;

Qy     125 GRVIPTFNVM---TDFCLHKIWVHENVDKYYVATDYVKEKLLEI-GTHPSNVKITGIPIR 180
           || :|   |:    ||    : |||::|       ||: :: : |  | ::
Db     427 GRYMPRMVVIPPGMDF--SNVVVHEDID----GDGDVKDDIVGLEGASPKSMP------- 473

Qy     181 PQFEESM-----PVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVV 235
           | : | |     |  |:    :    || :  :  | |  : ::||          : :
Db     474 PIWAEVMRFLTNPHKPMILALSRPDPKKNITTLVKAFGECRPLREL---------ANLTL 524

Qy     236 VCGKNTALKESLSALEAENGDKL-KVLGYVERID----------------ELFRITDCM 277
           : |     :: :  : | |   |  ||  ::: |                |::|:   |
Db     525 IMGN----RDDIDDMSAGNASVLTTVLKLIDKYDLYGSVAFPKHHNQADVPEIYRLAAKM 580
```

```
Qy    278 --------ITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEE-- 327
                  : :| |:|| || | |:|::  |     :  ||         ::|: |::
Db    581 KGVFINPALVEPFGLTLIEAAAHGLPIVATKNGGPVDITNA-----LNNGLLVDPHDQNA 635

Qy    328 ILESVTSLLADEDTLHRMKKN-IKDLHL 354
          | :::  |:||::     ::| ::::||
Db    636 IADALLKLVADKNLWQECRRNGLRNIHL 663


RESULT   4
US-09-051-341-7
; Sequence 7, Application US/09051341
; Patent No. 6124528
;  GENERAL INFORMATION:
;    APPLICANT:  Shewmaker, C. K.
;    TITLE OF INVENTION:  MODIFICATION OF SOLUBLE SOLIDS USING
;    TITLE OF INVENTION:  SUCROSE PHOSPHATE SYNTHASE ENCODING SEQUENCE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rae-Venter Law Group, P.C.
;      STREET:  260 Sheridan Avenue, Suite 440
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/051,341
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/17351
;      FILING DATE:  25-OCT-1996
;      APPLICATION NUMBER:  US 08/549,016
;      FILING DATE:  27-OCT-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/372,200
;      FILING DATE:  12-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Barbara Rae-Venter. Ph.D.,
;      REGISTRATION NUMBER:  32,750
;      REFERENCE/DOCKET NUMBER:  CGNE.110.02US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)328-4400
;      TELEFAX:  (415)328-4477
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1068 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-051-341-7


  Query Match             5.7%;  Score 112.5;  DB 4;  Length 1068;
  Best Local Similarity   20.9%;  Pred. No. 0.012;
  Matches   56;  Conservative   56;  Mismatches   87;  Indels   69;  Gaps   14;

Qy    125 GRVIPTFNVM---TDFCLHKIWVHENVDKYYVATDYVKEKLLEI-GTHPSNVKITGIPIR 180
          || :|   |:    ||    : |||::|        ||: :: : |  | ::
Db    427 GRYMPRMVVIPPGMDF--SNVVVHEDID----GDGDVKDDIVGLEGASPKSMP------- 473

Qy    181 PQFEESM-----PVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVV 235
          | : | |     |  |:    :    || :  :  | |  : ::||          : :
Db    474 PIWAEVMRFLTNPHKPMILALSRPDPKKNITTLVKAFGECRPLREL---------ANLTL 524

Qy    236 VCGKNTALKESLSALEAENGDKL-KVLGYVERID----------------ELFRITDCM 277
          : |     :: :  : | |   |  ||  ::: |                |::|:   |
Db    525 IMGN----RDDIDDMSAGNASVLTTVLKLIDKYDLYGSVAFPKHHNQADVPEIYRLAAKM 580

Qy    278 --------ITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEE-- 327
                  : :| |:|| || | |:|::  |     :  ||         ::|: |::
Db    581 KGVFINPALVEPFGLTLIEAAAHGLPIVATKNGGPVDITNA-----LNNGLLVDPHDQNA 635

Qy    328 ILESVTSLLADEDTLHRMKKN-IKDLHL 354
          | :::  |:||::     ::| ::::||
Db    636 IADALLKLVADKNLWQECRRNGLRNIHL 663


RESULT   5
PCT-US92-00282-7
; Sequence 7, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  529 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US92-00282-7


  Query Match             5.6%;  Score 110;  DB 5;  Length 529;
  Best Local Similarity   20.5%;  Pred. No. 0.0071;
  Matches   79;  Conservative   57;  Mismatches  137;  Indels  112;  Gaps   16;

Qy     40 SNLYQESNPIVSEVTQY--------------LYLK-----SFSIGKQFYRLF------- 72
          :| :  ||:|::: : :|              ||     ||    ||  ||
Db     95 NNHFAASSPLMAPLREYRNNMIVIDMCFFSCQSLLKDSATLSFLRENQFDALFTDPAMPC 154

Qy     73 ------YYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEYRRRTGR 126
                |  :  || |     |    : :|:           :  |:  || :  :
Db    155 GVILAEYLKLPSIYLFRG----FPCSLEHIGQ----------SPSPVSYVPRFYTKFSD 199

Qy    127 VIPTF---------NVMTDFCLHKIWVHENVDKYYV-ATDYVKEKLLEIGTHPSNVKITG 176
          : ||          |:: ::  | ::       || : |:| :|  :     | ::: :
Db    200 HM-TFPQRLANFIANILENYLYHCLY-----SKYEILASDLLKRDVSLPALHQNSLWLLR 253

Qy    177 IPIRPQFEESMPVGP---------IYKKYNLSPNKKVLLIMAGAHGVL------------ 215
              ||  -|| |           || |||   :  :  :| ||::
Db    254 YDF--VFEYPRPVMPNMIFIGGTNCKKKGNLSQEFEAYVNASGEHGIVVFSLGSMVSEIP 311
```

```
Qy     216 -KNVKELCENLVKDDQVQVVVVCG-------KNTALKESLSALEAENGDKLKVLGYVERI 267
           |   |: | | :  |  :    |       ||| | : |   :         :||: :
Db     312 EKKAMEIAEALGRIPQTLLWRYTGTRPSNLAKNTILVKWLPQND--------LLGHPK-- 361

Qy     268 DELFRITDCMITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEE 327
                  ||  |   : |    |||:::  |: | : :||   | |||  : :|  |
Db     362 ------ARAFITHSGSHGIYEGICNGVPMVM-MPLFGDQMDNAKRMETRGAGVTLNVLEM 414

Qy     328 ILESVTSLLADEDTLHRMKKNIKDL 352
             : : : |         |:||  |
Db     415 TADDLENALKTVINNKSYKENIMRL 439


RESULT   6
US-07-977-434-12
; Sequence 12, Application US/07977434
; Patent No. 5466591
;  GENERAL INFORMATION:
;    APPLICANT:  Gelfand, David H.
;    APPLICANT:  Abramson, Richard D.
;    TITLE OF INVENTION:  5' TO 3' EXONUCLEASE MUTATIONS OF
;    TITLE OF INVENTION:  THERMOSTABLE DNA POLYMERASES
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  New Jersey
;      ZIP:  07110-1199
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  7
;      SOFTWARE:  WordPerfect 2.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/977,434
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,490
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,466
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,213
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 523,394
;      FILING DATE:  15-MAY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 143,441
;      FILING DATE:  12-JAN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 063,509
;      FILING DATE:  17-JUN-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 899,241
;      FILING DATE:  22-AUG-1986
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 746,121
;      FILING DATE:  15-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US90/07641
;      FILING DATE:  21-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 585,471
;      FILING DATE:  20-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 455,611
;      FILING DATE:  22-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 609,157
;      FILING DATE:  02-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 557,517
;      FILING DATE:  24-JUL-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luann Cserr
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  Case No. 5466591 8753
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 814-2972
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  892 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-977-434-12


  Query Match             5.5%;  Score 108;  DB 1;  Length 892;
  Best Local Similarity   19.4%;  Pred. No. 0.026;
  Matches   72;  Conservative   63;  Mismatches  135;  Indels  102;  Gaps   15;

Qy      32 LGFQHVTVSNLYQESNPIVSEVTQYLYLK-----SFSIGKQFYRLFYYGVDKIYNKRKFN 86
           ||:: ::   |  |: |:       |: |:     |       ||:|
Db     431 LGYKMISFDELVNENVPLFGNDFSYVPLERAVEYSCEDADVTYRIF-------------- 476

Qy      87 IYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHE 146
                ::||  : |:: : :     |              |  :|:::  |: ::  |
Db     477 -------RKLGRKIYENEMEKLFYEIEM-------------PLIDVLSEMELNGVYFDE 515

Qy     147 NV-----DKYYVATDYVKEKLLEIGTHPSNVKITGIPIRPQFEESMPVGPI-YKKYNLSP 200
                 ||    | :|||: ||     |:             |  |  | ::| |::|
Db     516 EYLKELSKKYQEKMDGIKEKVFEIAGETFNL----------NSSTQVAYILFEKLNIAP 564

Qy     201 NKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVVVCGKNTALKESLSALEAENGDKLKV 260
            ||       | |      |: | | |: ::  ::              ||     |||
Db     565 YKKT------ATGKFSTNAEVLEELSKEHEIAKLL-------------LEYRKYQKLKS 604

Qy     261 LGYVE----------RIDELFRITDCMITKPGGITLTEATAIGVPVILYKPVPGQEKEN 309
             |::           |:   |  |    |  | :: :      :|    :   |:|
Db     605 T-YIDSIPLSINRKTNRVHTTFHQTG---TSTGRLSSSNPNLQNLPT---RSEEGKEIRK 657

Qy     310 ANFFEDRGAAIVVNRHEEI-LESVTSLLADEDTLHRMKKNIKDLHLANSSEVILEDILKE 368
           |   : :   |:   : :| |  :  :  ||: |   |::: |:|   :::     |
Db     658 AVRPQRQDWWILGADYSQIELRVLAHVSKDENLLKAFKEDL-DIHTITAAK-----IFGV 711

Qy     369 SEMMTAKQKAKV 380
           |||  ::|  :|
Db     712 SEMFVSEQMRRV 723


RESULT   7
US-08-458-819-12
; Sequence 12, Application US/08458819
; Patent No. 5795762
;  GENERAL INFORMATION:
;    APPLICANT:  Gelfand, David H.
;    APPLICANT:  Abramson, Richard D.
;    TITLE OF INVENTION:  5' TO 3' EXONUCLEASE MUTATIONS OF
;    TITLE OF INVENTION:  THERMOSTABLE DNA POLYMERASES
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  New Jersey
;      ZIP:  07110-1199
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  7
;      SOFTWARE:  WordPerfect 2.1
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,819
;      FILING DATE:  02-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/977,434
;      FILING DATE:  23-FEB-1993
;      APPLICATION NUMBER:  US 590,490
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,466
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,213
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 523,394
;      FILING DATE:  15-MAY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 143,441
;      FILING DATE:  12-JAN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 063,509
;      FILING DATE:  17-JUN-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 899,241
;      FILING DATE:  22-AUG-1986
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 746,121
;      FILING DATE:  15-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US90/07641
;      FILING DATE:  21-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 585,471
;      FILING DATE:  20-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 455,611
;      FILING DATE:  22-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 609,157
;      FILING DATE:  02-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 557,517
;      FILING DATE:  24-JUL-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Luann Cserr
;      REGISTRATION NUMBER:  31,822
;      REFERENCE/DOCKET NUMBER:  Case No. 5795762 8753
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 814-2972
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  892 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-458-819-12


  Query Match             5.5%;  Score 108;  DB 1;  Length 892;
  Best Local Similarity   19.4%;  Pred. No. 0.026;
  Matches   72;  Conservative   63;  Mismatches  135;  Indels  102;  Gaps   15;

Qy      32 LGFQHVTVSNLYQESNPIVSEVTQYLYLK-----SFSIGKQFYRLFYYGVDKIYNKRKFN 86
           ||:: ::   |  |: |:      |: |:     |       ||:|
Db     431 LGYKMISFDELVNENVPLFGNDFSYVPLERAVEYSCEDADVTYRIF-------------- 476

Qy      87 IYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHE 146
             ::||  : |:: : :     |             |  :|:::  |: ::  |
Db     477 -------RKLGRKIYENEMEKLFYEIEM-------------PLIDVLSEMELNGVYFDE 515

Qy     147 NV-----DKYYVATDYVKEKLLEIGTHPSNVKITGIPIRPQFEESMPVGPI-YKKYNLSP 200
                ||    | :|||: ||     |:           |  |  | ::| |::|
Db     516 EYLKELSKKYQEKMDGIKEKVFEIAGETFNL----------NSSTQVAYILFEKLNIAP 564

Qy     201 NKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVVVCGKNTALKESLSALEAENGDKLKV 260
           ||       | |       |: | | |: ::  ::              ||     |||
Db     565 YKKT------ATGKFSTNAEVLEELSKEHEIAKLL-------------LEYRKYQKLKS 604

Qy     261 LGYVE----------RIDELFRITDCMITKPGGITLTEATAIGVPVILYKPVPGQEKEN 309
             |::          |:   |  |    |  | :: :      :|    :   |:|
Db     605 T-YIDSIPLSINRKTNRVHTTFHQTG---TSTGRLSSSNPNLQNLPT---RSEEGKEIRK 657

Qy     310 ANFFEDRGAAIVVNRHEEI-LESVTSLLADEDTLHRMKKNIKDLHLANSSEVILEDILKE 368
           |   : :   |:   : :| |  :  :  ||: |   |::: |:|   :::     |
Db     658 AVRPQRQDWWILGADYSQIELRVLAHVSKDENLLKAFKEDL-DIHTITAAK-----IFGV 711

Qy     369 SEMMTAKQKAKV 380
           |||  ::|  :|
Db     712 SEMFVSEQMRRV 723


RESULT   8
PCT-US91-07035-12
; Sequence 12, Application PC/TUS9107035
;  GENERAL INFORMATION:
;    APPLICANT:  Gelfand, David H.
;    APPLICANT:  Abramson, Richard D.
;    TITLE OF INVENTION:  5' TO 3' EXONUCLEASE MUTATIONS OF
;    TITLE OF INVENTION:  THERMOSTABLE DNA POLYMERASES
;    NUMBER OF SEQUENCES:  38
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cetus Corporation
;      STREET:  1400 Fifty-third Street
;      CITY:  Emeryville
;      STATE:  California
;      ZIP:  94608
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US91/07035
;      FILING DATE:  19910930
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,490
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,466
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 590,213
;      FILING DATE:  28-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 523,394
;      FILING DATE:  15-MAY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 143,441
;      FILING DATE:  12-JAN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 063,509
;      FILING DATE:  17-JUN-1987
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 899,241
;      FILING DATE:  22-AUG-1986
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 746,121
;      FILING DATE:  15-AUG-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/US90/07641
;      FILING DATE:  21-DEC-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 585,471
```

```
;      FILING DATE:  20-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 455,611
;      FILING DATE:  22-DEC-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 609,157
;      FILING DATE:  02-NOV-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 557,517
;      FILING DATE:  24-JUL-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sias Ph.D, Stacey R.
;      REGISTRATION NUMBER:  32,630
;      REFERENCE/DOCKET NUMBER:  Case No. 2580
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-420-3300
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  892 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US91-07035-12


  Query Match             5.5%;  Score 108;  DB 5;  Length 892;
  Best Local Similarity   19.4%;  Pred. No. 0.026;
  Matches   72;  Conservative   63;  Mismatches  135;  Indels  102;  Gaps   15;

Qy      32 LGFQHVTVSNLYQESNPIVSEVTQYLYLK-----SFSIGKQFYRLFYYGVDKIYNKRKFN 86
           ||:: ::   |  |: |:      |: ||      |       ||:|
Db     431 LGYKMISFDELVNENVPLFGNDFSYVPLERAVEYSCEDADVTYRIF-------------- 476

Qy      87 IYFKMGNKRLGELVDEHQPDIIINTFPMIVVPEYRRRTGRVIPTFNVMTDFCLHKIWVHE 146
             ::||  : |:: : :     |              |  :|:::  |: ::  |
Db     477 -------RKLGRKIYENEMEKLFYEIEM-------------PLIDVLSEMELNGVYFDE 515

Qy     147 NV-----DKYYVATDYVKEKLLEIGTHPSNVKITGIPIRPQFEESMPVGPI-YKKYNLSP 200
             ||     | :|||: ||     |:             |  |  | ::| |::|
Db     516 EYLKELSKKYQEKMDGIKEKVFEIAGETFNL----------NSSTQVAYILFEKLNIAP 564

Qy     201 NKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVVVCGKNTALKESLSALEAENGDKLKV 260
            ||       | |       |: | |  |: ::  ::             ||     |||
Db     565 YKKT------ATGKFSTNAEVLEELSKEHEIAKLL-------------LEYRKYQKLKS 604

Qy     261 LGYVE-----------RIDELFRITDCMITKPGGITLTEATAIGVPVILYKPVPGQEKEN 309
             |::           |:  |  |    |  | :: :      :|    :   |:|
Db     605 T-YIDSIPLSINRKTNRVHTTFHQTG---TSTGRLSSSNPNLQNLPT---RSEEGKEIRK 657

Qy     310 ANFFEDRGAAIVVNRHEEI-LESVTSLLADEDTLHRMKKNIKDLHLANSSEVILEDILKE 368
           |   : :   |:   : :| | :  :  ||: |   |::: |:|   :::     |
Db     658 AVRPQRQDWWILGADYSQIELRVLAHVSKDENLLKAFKEDL-DIHTITAAK-----IFGV 711

Qy     369 SEMMTAKQKAKV 380
           |||  ::|  :|
Db     712 SEMFVSEQMRRV 723


RESULT  9
US-08-843-530B-36
; Sequence 36, Application US/08843530B
; Patent No. 5939306
;  GENERAL INFORMATION:
;    APPLICANT:  Selitrennikoff, Claude
;    APPLICANT:  Agnan, Jacqueline
;    APPLICANT:  Alex, Lisa A.
;    APPLICANT:  Simon, Melvin I.
;    TITLE OF INVENTION:  Osmosensing Histidine Kinases
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Medlen & Carroll, LLP
;      STREET:  220 Montgomery Street, Suite 2200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  United States of America
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/843,530B
;      FILING DATE:  16-APR-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MacKnight, Kamrin T.
;      REGISTRATION NUMBER:  38,230
;      REFERENCE/DOCKET NUMBER:  UTC-02717
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1220 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  protein
US-08-843-530B-36


  Query Match             5.4%;  Score 105.5;  DB 2;  Length 1220;
  Best Local Similarity   20.2%;  Pred. No. 0.076;
  Matches   76;  Conservative   60;  Mismatches  124;  Indels  117;  Gaps   19;

Qy      11 TANYGN---GHVQVAKTLYEQCVRLGFQHVTVSNLYQES------NPIVSEVTQYL---- 57
           |:|| |     : :|  |    :       |:: || ::      : : | :| |:
Db      47 TSNYKNLRSDRLYIAAQLKSSQI-----DQTLNYLYYQAYYLASRDALQSSLTSYVAGNK 101

Qy      58 ----YLKSFSIGKQFY---RLFYYGVDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIIN 110
                :: | |: ::|     |||  | |:|:   ||      |   |:|: |   |:: :
Db     102 SADNWVDSLSVIQKFLSSSNLFY--VAKVYDS-SFNAVLNATNNGTGDLIPE---DVLDS 155

Qy     111 TFPMIVVPEYRRRTGRVIP----TFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIG 166
            ||:          |   :|    |  ::||  .|:             :|||:    | |
Db     156 LFPL--------STDTPLPSSLETIGILTDPVLN------------STDYLMSMSLPIF 194

Qy     167 THPS------------NVKITGIPIRPQFE-----ESMPVGPIYKKYNLSPNKKVLLIMA 209
             :||            : ::   ::  |      |   :  |   ||          :
Db     195 ANPSIILTDSRVYGYITIIMSAEGLKSVFNDTTALEHSTIAIISAVYNSQGKASGYHFVF 254

Qy     210 GAHGVLKNVKELCENLVKDDQVQVVVVCGKNTALKESLSALEAENGDKLKVLGYVERIDE 269
             :|   :: :   ::  |  :       ||  :||:: : |   |      |||
Db     255 PPYGSRSDLPQKVFSIKNDTFISSAFRNGKGGSLKQT-NILSTRN----TALGY------ 303

Qy     270 LFRITDC---------MITKPGGITLTEAT-----------AIGVPVIL---------YK 300
           : |          ::::|  : |: ||           |||| |||          :
Db     304 ----SPCSFNLVNWVAIVSQPESVFLSPATKLAKIITGTVIAIGVFVILLTLPLAHWAVQ 359

Qy     301 PVPGQEKENANFFEDRG 317
           |:   :|      | ||
Db     360 PIVRLQKATELITEGRG 376


RESULT  10
US-09-007-476-2
; Sequence 2, Application US/09007476
; Patent No. 6159949
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Michael T.
;    TITLE OF INVENTION:  No. 6159949el FtsY
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  4000 Bell Atlantic Tower, 1717 Arch Stre
```

```
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  US
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/007,476
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Falk, Stephen T
;      REGISTRATION NUMBER:  36,795
;      REFERENCE/DOCKET NUMBER:  GM10079
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-994-2488
;      TELEFAX:  215-994-2222
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  416 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-007-476-2


  Query Match             5.2%;  Score 102.5;  DB 4;  Length 416;
  Best Local Similarity   24.5%;  Pred. No. 0.029;
  Matches   45;  Conservative   35;  Mismatches   83;  Indels   21;  Gaps    7;

Qy     176 GIPIRPQFEESMPVGPIYKKYNLSPNKKVLLIMAGAHGVLKNVKELCENLVKDDQVQVVV 235
           |: :  | | | |   :|   | : || | :::    | |:|   | : |   ::|: |:
Db     241 GVDVISQSEGSDPAAVMYDAINAAKNKGVDILICDTAGRLQNKTNLMQEL---EKVKRVI 297

Qy     236 VCGKNTALKESLSALEAENGDKLKVLGYVERIDELFRITDCMITK-----PGGITLTEAT 290
             |  |:|  |:|  |     |       |:  :|  ::||      |||  |
Db     298 NRAVPDAPHEALLCLDATTGQ--NALSQARNFKEVTNVTGIVLTKLDGTAKGGIVLAIRN 355

Qy     291 AIGVPVILYKPVP-GQEKENANFFEDRG-----AAIVVNRHEEI--LESVTSLLADEDTL 342
           : :||   | |  |:: ::   |          | :: :::||  :|:   :  ::|
Db     356 ELHIPV---KYVGLGEQLDDLQPFNPESYVYGLFADMIEQNDEITTVENDQIVTEEKDDN 412

Qy     343 HRMK 346
           |  |
Db     413 HGAK 416


RESULT  11
US-08-533-669A-10
; Sequence 10, Application US/08533669A
; Patent No. 5834592
;  GENERAL INFORMATION:
;    APPLICANT:  Corixa Corporation
;    TITLE OF INVENTION:  LEISHMANIA ANTIGENS FOR USE IN THE
;    TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF LEISHMANIASIS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/533,669A
;      FILING DATE:  22-SEP-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  403 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-533-669A-10


  Query Match             5.1%;  Score 101;  DB 2;  Length 403;
  Best Local Similarity   20.8%;  Pred. No. 0.04;
  Matches   61;  Conservative   50;  Mismatches  110;  Indels   72;  Gaps   12;

Qy     126 RVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLL-----------------EIGT 167
           | ||:|: |          :|:|: :   :  : |   :                 : ||
Db      23 RPIPSFDDMP--------LHQNLLRGIYSYGFEKPSSIQQRAIAPFTRGGDIIAQAQSGT 74

Qy     168 HPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLL----IMAGAHGVLKNVKELCE 223
             :     |:  |  |  ::  | :     ||| ::: |    :::     | |  : ||
Db      75 GKTGAFSIGLLQRLDFRHNLIQGLV-----LSPTRELALQTAEVISRIGEFLSNSAKFCE 129

Qy     224 NL-----VKDD----QVQVVVVCG---------KNTALK-ESLSALEAENGDKLKVLGYV 264
                  |:||    |  |||  |          |  ||: |||  |  :  |::   |:
Db     130 TFVGGTRVQDDLRKLQAGVVVAVGTPGRVSDVIKRGALRTESLRVLVLDEADEMLSQGFA 189

Qy     265 ERIDELFRITDCMITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNR 324
           ::| |:||     : |   : |  ||           :| :  |    |       |:| |
Db     190 DQIYEIFR----FLPKDIQVALFSAT----------MPEEVLELTKKFMRDPVRILVKR 234

Qy     325 HEEILESVTSLLADEDTLHRMKKNIKDLH--LANSSEVILEDILKESEMMTAK 375
           || :     :  |::   : ||:  :: :  ||  :  :: : :  |
Db     235 ESLTLEGIKQFFIAVEEEHKL-DTLMDLYETVSIAQSVIFANTRRKVDWIAEK 286


RESULT  12
US-08-607-509-2
; Sequence 2, Application US/08607509
; Patent No. 5876735
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  METHODS FOR ENHANCEMENT OF PROTECTIVE IMMUNE RESPONSES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/607,509
;      FILING DATE:  16-FEB-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
```

```
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.404C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  403 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-607-509-2


  Query Match             5.1%;  Score 101;  DB 2;  Length 403;
  Best Local Similarity   20.8%;  Pred. No. 0.04;
  Matches   61;  Conservative   50;  Mismatches  110;  Indels   72;  Gaps   12;

Qy     126 RVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLL-----------------EIGT 167
           | ||:|: |         :|:|: :   :  : |   :                 : ||
Db      23 RPIPSFDDMP--------LHQNLLRGIYSYGFEKPSSIQQRAIAPFTRGGDIIAQAQSGT 74

Qy     168 HPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLL----IMAGAHGVLKNVKELCE 223
             :     |:  |  |  ::  | :     ||| ::: |    :::     | |  : ||
Db      75 GKTGAFSIGLLQRLDFRHNLIQGLV-----LSPTRELALQTAEVISRIGEFLSNSAKFCE 129

Qy     224 NL-----VKDD----QVQVVVVCG---------KNTALK-ESLSALEAENGDKLKVLGYV 264
                 |:||    |  |||  |          |  ||: |||  |  :  |::   |:
Db     130 TFVGGTRVQDDLRKLQAGVVVAVGTPGRVSDVIKRGALRTESLRVLVLDEADEMLSQGFA 189

Qy     265 ERIDELFRITDCMITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNR 324
           ::| |:||     : |   : |  ||           :| :  |    |      |:| |
Db     190 DQIYEIFR----FLPKDIQVALFSAT----------MPEEVLELTKKFMRDPVRILVKR 234

Qy     325 HEEILESVTSLLADEDTLHRMKKNIKDLH--LANSSEVILEDILKESEMMTAK 375
             || :       :  |::   : ||:  :: :  ||  :  :: : :  |
Db     235 ESLTLEGIKQFFIAVEEEHKL-DTLMDLYETVSIAQSVIFANTRRKVDWIAEK 286


RESULT  13
US-08-607-509-4
; Sequence 4, Application US/08607509
; Patent No. 5876735
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  METHODS FOR ENHANCEMENT OF PROTECTIVE IMMUNE RESPONSES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/607,509
;      FILING DATE:  16-FEB-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.404C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  403 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-607-509-4


  Query Match             5.1%;  Score 101;  DB 2;  Length 403;
  Best Local Similarity   20.5%;  Pred. No. 0.04;
  Matches   60;  Conservative   51;  Mismatches  110;  Indels   72;  Gaps   12;

Qy     126 RVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLL-----------------EIGT 167
           | ||:|: |         :|:|: :   :  : |   :                 : ||
Db      23 RPIPSFDDMP--------LHQNLLRGIYSYGFEKPSSIQQRAIAPFTRGGDIIAQAQSGT 74

Qy     168 HPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLL----IMAGAHGVLKNVKELCE 223
             :     |:  |  |  ::  | :     ||| ::: |    :::     | |  : ||
Db      75 GKTGAFSIGLLQRLDFRHNLIQGLV-----LSPTRELALQTAEVISRIGEFLSNSSKFCE 129

Qy     224 NL-----VKDD----QVQVVVVCG---------KNTALK-ESLSALEAENGDKLKVLGYV 264
                 |:||    |  |:|  |          |  ||: |||  |  :  |::   |:
Db     130 TFVGGTRVQDDLRKLQAGVIVAVGTPGRVSDVIKRGALRTESLRVLVLDEADEMLSQGFA 189

Qy     265 ERIDELFRITDCMITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNR 324
           ::| |:||     : |   : |  ||           :| :  |    |      |:| |
Db     190 DQIYEIFR----FLPKDIQVALFSAT----------MPEEVLELTKKFMRDPVRILVKR 234

Qy     325 HEEILESVTSLLADEDTLHRMKKNIKDLH--LANSSEVILEDILKESEMMTAK 375
             || :       :  |::   : ||:  :: :  ||  :  :: : :  |
Db     235 ESLTLEGIKQFFIAVEEEHKL-DTLMDLYETVSIAQSVIFANTRRKVDWIAEK 286


RESULT  14
US-08-454-036-2
; Sequence 2, Application US/08454036
; Patent No. 5876966
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE STIMULATION AND ENHANCEMENT O
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/454,036
;      FILING DATE:  30-MAY-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kadlecek, Ann T.
;      REGISTRATION NUMBER:  P-39,244
;      REFERENCE/DOCKET NUMBER:  210121.404C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;      TELEX:  3723836 SEEDANDBERRY
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  403 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-454-036-2


  Query Match             5.1%;  Score 101;  DB 2;  Length 403;
```

```
  Best Local Similarity   20.8%;  Pred. No. 0.04;
  Matches   61;  Conservative   50;  Mismatches  110;  Indels   72;  Gaps   12;

Qy     126 RVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLL-----------------EIGT 167
           | ||:|: |         :|:|: :   :  : |   :                  : ||
Db      23 RPIPSFDDMP--------LHQNLLRGIYSYGFEKPSSIQQRAIAPFTRGGDIIAQAQSGT 74

Qy     168 HPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLL----IMAGAHGVLKNVKELCE 223
           :    |:  |  |  ::  | :     ||| ::: |    :::     | |  : ||
Db      75 GKTGAFSIGLLQRLDFRHNLIQGLV-----LSPTRELALQTAEVISRIGEFLSNSAKFCE 129

Qy     224 NL-----VKDD----QVQVVVVCG---------KNTALK-ESLSALEAENGDKLKVLGYV 264
                 |:||    |  |||  |          |  ||: |||  |  :  |::   |:
Db     130 TFVGGTRVQDDLRKLQAGVVVAVGTPGRVSDVIKRGALRTESLRVLVLDEADEMLSQGFA 189

Qy     265 ERIDELFRITDCMITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNR 324
           ::| |:||     : |   : |  ||           :| :  |    |      |:| |
Db     190 DQIYEIFR----FLPKDIQVALFSAT----------MPEEVLELTKKFMRDPVRILVKR 234

Qy     325 HEEILESVTSLLADEDTLHRMKKNIKDLH--LANSSEVILEDILKESEMMTAK 375
             || :       :  |::   : ||:  :: :  ||  :  :: : :  |
Db     235 ESLTLEGIKQFFIAVEEEHKL-DTLMDLYETVSIAQSVIFANTRRKVDWIAEK 286


RESULT  15
US-08-634-642-2
; Sequence 2, Application US/08634642
; Patent No. 5879687
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  METHODS FOR ENHANCEMENT OF
;    TITLE OF INVENTION:  PROTECTIVE IMMUNE RESPONSES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,642
;      FILING DATE:  18-APR-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210131.404C4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  403 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-634-642-2


  Query Match             5.1%;  Score 101;  DB 2;  Length 403;
  Best Local Similarity   20.8%;  Pred. No. 0.04;
  Matches   61;  Conservative   50;  Mismatches  110;  Indels   72;  Gaps   12;

Qy     126 RVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLL-----------------EIGT 167
           | ||:|: |         :|:|: :   :  : |   :                  : ||
Db      23 RPIPSFDDMP--------LHQNLLRGIYSYGFEKPSSIQQRAIAPFTRGGDIIAQAQSGT 74

Qy     168 HPSNVKITGIPIRPQFEESMPVGPIYKKYNLSPNKKVLL----IMAGAHGVLKNVKELCE 223
           :    |:  |  |  ::  | :     ||| ::: |    :::     | |  : ||
Db      75 GKTGAFSIGLLQRLDFRHNLIQGLV-----LSPTRELALQTAEVISRIGEFLSNSAKFCE 129

Qy     224 NL-----VKDD----QVQVVVVCG---------KNTALK-ESLSALEAENGDKLKVLGYV 264
                 |:||    |  |||  |          |  ||: |||  |  :  |::   |:
Db     130 TFVGGTRVQDDLRKLQAGVVVAVGTPGRVSDVIKRGALRTESLRVLVLDEADEMLSQGFA 189

Qy     265 ERIDELFRITDCMITKPGGITLTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNR 324
           ::| |:||     : |   : |  ||           :| :  |    |      |:| |
Db     190 DQIYEIFR----FLPKDIQVALFSAT----------MPEEVLELTKKFMRDPVRILVKR 234

Qy     325 HEEILESVTSLLADEDTLHRMKKNIKDLH--LANSSEVILEDILKESEMMTAK 375
             || :       :  |::   : ||:  :: :  ||  :  :: : :  |
Db     235 ESLTLEGIKQFFIAVEEEHKL-DTLMDLYETVSIAQSVIFANTRRKVDWIAEK 286


Search completed: June 29, 2001, 08:59:39
Job time: 177 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 11:53:36 ; Search time 2892.01 Seconds
(without alignments)
5214.731 Million cell updates/sec

Title: US-09-668-788-3
Perfect score: 975
Sequence: 1 atggttactcaaaataaaaa.....gtaaaacgctgatactccag 975

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	975	100.0	975	9	AX016297	AX016297 Sequence
	2	964	98.9	7791	3	SAY14370	Y14370 Staphylococ
	3	636.2	65.3	3118	2	AF270166	AF270166 Staphyloc
	4	636.2	65.3	3493	2	AF270394	AF270394 Staphyloc
c	5	437	44.8	2600	9	AR106079	AR106079 Sequence
c	6	437	44.8	2600	10	E35634	E35634 MurE. 2/200
	7	118	12.1	1149	9	AX016296	AX016296 Sequence
c	8	118	12.1	35040	2	BACYACA	L77246 Bacillus su

	9	118	12.1	213680	2	BSUB0012	Z99115 Bacillus su
	10	72	7.4	242893	83	CEY53C12	Z92859 Caenorhabdi
	11	60.2	6.2	1496	6	CEY53C12D	AL033123 Caenorhab
	12	57.4	5.9	1648	13	ATMGDGSYN	AJ000331 Arabidops
	13	56.8	5.8	44352	6	CELH28G03	AF098501 Caenorhab
	14	56.4	5.8	152878	6	CEY18D10A	AL034393 Caenorhab
	15	56.4	5.8	247332	83	CEY18D10	AL008871 Caenorhab
c	16	55.4	5.7	148061	87	AC010727	AC010727 Homo sapi
	17	55	5.6	1574	12	AB047398	AB047398 Arabidops
	18	54.6	5.6	1141	10	AX083744	AX083744 Sequence
c	19	54.4	5.6	175415	82	AL590063	AL590063 Homo sapi
c	20	54.2	5.6	169437	81	AL391538	AL391538 Homo sapi
c	21	54	5.5	214631	72	AC060805	AC060805 Homo sapi
c	22	53.6	5.5	53932	68	AC023371	AC023371 Homo sapi
	23	52.8	5.4	54345	4	AC084152	AC084152 Caenorhab
	24	52.6	5.4	1154	5	AF044864	AF044864 Pieris ra
c	25	52.6	5.4	1189	80	AL359196	AL359196 Homo sapi
c	26	52.4	5.4	867	53	CNS075BG	AL429890 clone BA0
c	27	52	5.3	205429	60	AC005506	AC005506 Plasmodiu
c	28	51.4	5.3	129388	69	AC025184	AC025184 Homo sapi
	29	51.2	5.3	1775	12	AB047399	AB047399 Arabidops
	30	51.2	5.3	1929	13	AF241797	AF241797 Arabidops
c	31	51	5.2	1038	53	CNS06L7M	AL403832 T3 end of
	32	51	5.2	186625	86	AC006487	AC006487 Homo sapi
c	33	50.8	5.2	155317	73	AC067973	AC067973 Homo sapi
	34	50.4	5.2	100925	87	AC012627	AC012627 Homo sapi
	35	50.4	5.2	170432	67	AC022145	AC022145 Homo sapi
c	36	50.4	5.2	235532	87	AC008739	AC008739 Homo sapi
	37	50.2	5.1	910	53	CNS01G8P	AL142826 Anopheles
	38	50.2	5.1	156060	60	AC004153	AC004153 Plasmodiu
c	39	50	5.1	1496	6	CEY53C12D	AL033123 Caenorhab
	40	50	5.1	149752	85	AC004616	AC004616 Homo sapi
c	41	50	5.1	242893	83	CEY53C12	Z92859 Caenorhabdi
	42	49.8	5.1	759	53	CNS06QXV	AL411257 T7 end of
c	43	49.8	5.1	1013	53	CNS06RPQ	AL412260 T7 end of
	44	49.8	5.1	166331	80	AL360177	AL360177 Homo sapi
c	45	49.8	5.1	171863	91	AP000459	AP000459 Homo sapi

ALIGNMENTS

```
RESULT   1
AX016297
LOCUS       AX016297      975 bp    DNA             PAT       07-SEP-2000
DEFINITION  Sequence 2 from Patent WO9949052.
ACCESSION   AX016297
VERSION     AX016297.1  GI:10041860
KEYWORDS    .
SOURCE      Staphylococcus aureus.
  ORGANISM  Staphylococcus aureus
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE   1  (bases 1 to 975)
  AUTHORS   Zaehringer,U., Heinz,E., Jorasch,P. and Wolter,F.P.
  TITLE     Processive glycosyltransferase
  JOURNAL   Patent: WO 9949052-A 2 30-SEP-1999;
            FORSCHUNGSZENTRUM BORSTEL (DE); ZAEHRINGER ULRICH (DE); HEINZ ERNST
            (DE); JORASCH PETRA (DE); WOLTER FRANK P (DE); GVS GES FUER ERWERB
            UND VERWER (DE)
FEATURES             Location/Qualifiers
     source          1. .975
                     /organism="Staphylococcus aureus"
                     /db_xref="taxon:1280"
BASE COUNT      350 a    152 c    169 g    304 t
ORIGIN
```

Query Match 100.0%; Score 975; DB 9; Length 975;
Best Local Similarity 100.0%; Pred. No. 1.5e-164;
Matches 975; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 atggttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGTTACTCAAAATAAAAAGATATTGATTATTACTGGCTCATTCGGTAACGGTCATATG 60

Qy     61 caagttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CAAGTTACACAGAGTATCGTTAATCAACTTAATGATATGAATCTAGACCATTTAAGCGTC 120

Qy    121 attgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatgg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATTGAGCACGATTTATTTATGGAAGCTCATCCAATTTTGACTTCTATTTGTAAAAAATGG 180

Qy    181 tatatcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgccca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TATATCAATAGCTTTAAATATTTTAGAAATATGTACAAAGGGTTTTATTACAGCCGCCCA 240

Qy    241 gataaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GATAAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTTATTG 300

Qy    301 ataaaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtacta 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATAAAAGAAAAGCCAGATTTAATATTATTAACGTTTCCTACACCAGTTATGTCGGTACTA 360

Qy    361 actgagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ACTGAGCAATTTAACATTAATATTCCAGTTGCTACAGTGATGACAGACTATCGCTTACAT 420

Qy    421 aaaaactggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaa 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAAAACTGGATTACGCCGTATTCAACAAGATATTATGTGGCAACAAAAGAAACGAAACAA 480

Qy    481 gacttcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgat 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GACTTCATAGACGTAGGTATTGATCCTTCAACAGTTAAAGTGACAGGTATTCCTATTGAT 540

Qy    541 aacaaatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AACAAATTTGAAACGCCTATTAATCAAAAGCAGTGGTTAATAGACAACAACTTAGATCCA 600

Qy    601 gataagcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacg 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GATAAGCAAACTATTTTAATGTCAGCTGGTGCATTTGGTGTATCTAAAGGTTTTGACACG 660

Qy    661 atgattactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaag 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATGATTACTGATATATTAGCGAAAAGTGCAAATGCACAAGTAGTTATGATTTGTGGTAAG 720

Qy    721 agcaaagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgatt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 AGCAAAGAGCTAAAGCGTTCTTTAACAGCTAAGTTTAAATTAACGAGAATGTATTTGATT 780

Qy    781 ctaggttataccaaacacatgaatgaatggatggcatcaagtcaacttatgattacgaaa 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CTAGGTTATACCAAACACATGAATGAATGGATGGCATCAAGTCAACTTATGATTACGAAA 840

Qy    841 cctggtggtatcacaataactgaaggtttcgcccgttgtattccaatgattttcctaaat 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CCTGGTGGTATCACAATAACTGAAGGTTTCGCCCGTTGTATTCCAATGATTTTCCTAAAT 900

Qy    901 cctgcacctggtcaagagcttgaaaatgccttttacttgaagaaaaaggttttggtaaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CCTGCACCTGGTCAAGAGCTTGAAAATGCCTTTTACTTTGAAGAAAAAGGTTTTGGTAAA 960

Qy    961 acgctgatactccag 975
          ||||||||||||||||
Db    961 ACGCTGATACTCCAG 975
```

```
RESULT   2
SAY14370
LOCUS       SAY14370      7791 bp    DNA             BCT       24-JUN-1998
DEFINITION  Staphylococcus aureus RF3, murE, ypfP genes.
ACCESSION   Y14370
```

```
VERSION     Y14370.1  GI:3256221
KEYWORDS    murE gene; peptide chain release factor 3; RF3 gene;
            UDP-N-acetylmuramyl-tripeptide synthetase; ypfP gene.
SOURCE      Staphylococcus aureus.
  ORGANISM  Staphylococcus aureus
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE   1  (bases 1 to 7791)
  AUTHORS   Ludovice,A.M., Wu,S. and de Lencastre,H.
  TITLE     Molecular cloning and DNA sequencing of the Staphylococcus aureus
            UDP-N-acetylmuramyl tripeptide synthetase (murE) gene, essential
            for the optimal expression of methicillin resistance
  JOURNAL   Microb. Drug Res. 4, 85-90 (1998)
REFERENCE   2  (bases 1 to 7791)
  AUTHORS   Ludovice,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1997) A.M. Ludovice, Instituto de Tecnologia
            Qiomica e Biologica, Universidade Nova de Lisbon, R. da Quinta
            Grande 6, Apartado 127, 2780 Oeiras, PORTUGAL
  REMARK    Revised by author 16-JAN-98
COMMENT     Related sequence: L77246.
FEATURES             Location/Qualifiers
     source          1. .7791
                     /organism="Staphylococcus aureus"
                     /strain="COL"
                     /db_xref="taxon:1280"
     gene            complement(1435. .3114)
                     /gene="RF3"
     CDS             complement(1435. .3000)
                     /gene="RF3"
                     /codon_start=1
                     /transl_table=11
                     /product="peptide chain release factor 3"
                     /protein_id="CAA74739.1"
                     /db_xref="GI:3256222"
                     /db_xref="SPTREMBL:O86490"
                     /translation="MNLKQEVESRKTFAIISHPDAGKTTLTEKLLYFSGAIREAGTVK
                     GKKLVNLRQVTWMKVEQERGISVTSSVMQFDYDDYEINILDTPGHEDFSEDTYRTLMA
                     VDSAVMVIDCAKGVEPPTLKLFKVCKMRGIPIFTFINKLDRVGKEPFELLDEIEETLN
                     IETYPMNWPIGMGQSFFGIIDRKSKTIEPFRDEENILHLNDDFELEEDHAITNDSDFE
                     QAIEELMLVEEAGEAFDNDALLSGDLTPVFFGSALANFGVQNFLNAYVDFAPMPNARQ
                     TKENVEVSPFDDSFSGFIFKIQANMDPKHRDRIAFMRVVSGAFERVWMLLCNVLIKSK
                     RSHVQRHLWQTIKKLVNHAVAGDIIGLYDTGNYQIGDTLVGGKQTYSFQDLPQFTPEI
                     FMKVSAKNVMKQKHFHKGIEQLVQEGAIQYYKTLHTNQIILGAVGQLQFEVFEHRMKN
                     EYNVDVVMEPVGRKIARWDIENEDQITDKMNTSRSILVKDRYDDLVFLFENEFATRWF
                     EEKFPEIKLYSLL"
     RBS             complement(3009. .3013)
                     /gene="RF3"
     -10_signal      complement(3085. .3091)
                     /gene="RF3"
     -35_signal      complement(3109. .3114)
                     /gene="RF3"
     gene            complement(3244. .4784)
                     /gene="murE"
     CDS             complement(3244. .4722)
                     /gene="murE"
                     /codon_start=1
                     /transl_table=11
                     /product="UDP-N-acetylmuramyl-tripeptide synthetase"
                     /protein_id="CAA74740.1"
                     /db_xref="GI:3256223"
                     /db_xref="SPTREMBL:O86491"
                     /translation="MQVRCLKKVKVKRVLGSLEQQIDDITTDSRTAREGSIFVASVGY
                     TVDSHKFCQNVADQGCKLVVVNKEQSLPANVTQVVVPDTLRVASILAHTLYDYPSHQL
                     VTFGVTGTNGKTSIATMIHLIQRKLQKNSAYLGTNGFQINETKTKGANTTPETVSLTK
                     KIKEAVDAGAESMTLEVSSHGLVLGRLRSVEFDVAIFSNLTQDHLDFHGTMEAYGHAK
                     SLLFSQLGEDLSKEKYVVLNNDDSFSEYLRTVTPYEVFSYGIDEEAQFMAKNIQESLQ
                     GVSFDFVTPFGTYPVKSPYVGKFNISNIMAAMIAVWSKGTSLETIIKAVENLEPVEGR
                     LEVLDPSLPIDLIIDYAHTADGMNKLIDAVQPFVKQKLIFLVGMAGERDLTKTPEMGR
                     VACRADYVIFTPDNPANDDPKMLTAELAKGATHQNYIEFDDRAEGIKHAIDIAEPGDT
                     VVLASKGREPYQIMPGHIKVPHRDDLIGLEAAYKKFGGGPVD"
     RBS             complement(4736. .4740)
                     /gene="murE"
     -10_signal      complement(4758. .4763)
                     /gene="murE"
     -35_signal      complement(4779. .4784)
                     /gene="murE"
     gene            5086. .6335
                     /gene="ypfP"
     -35_signal      5086. .5092
                     /gene="ypfP"
     -10_signal      5107. .5112
                     /gene="ypfP"
     RBS             5145. .5150
                     /gene="ypfP"
     CDS             5160. .6335
                     /gene="ypfP"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA74741.1"
                     /db_xref="GI:3256224"
                     /db_xref="SPTREMBL:O86492"
                     /translation="MVTQNKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHD
                     LFMEAHPILTSICKKWYINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIK
                     EKPDLILLTFPTPVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQ
                     DFIDVGIDPSTVKVTGIPIDNKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGF
                     DTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQL
                     MITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVAS
                     LTNGNEQLTNMISTMEQDKIKYATQTICRDLLDLIGHSSQPQEIYGKVPLYARFFVK"
BASE COUNT     2517 a   1452 c   1130 g   2668 t     24 others
ORIGIN
```

```
  Query Match             98.9%;  Score 964;  DB 3;  Length 7791;
  Best Local Similarity   99.9%;  Pred. No. 1.2e-162;
  Matches  975;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy       1 atggttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatg 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5160 ATGGTTACTCAAAATAAAAAGATATTGATTATTACTGGCTCATTCGGTAACGGTCATATG 5219

Qy      61 caagttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5220 CAAGTTACACAGAGTATCGTTAATCAACTTAATGATATGAATCTAGACCATTTAAGCGTC 5279

Qy     121 attgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatgg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5280 ATTGAGCACGATTTATTTATGGAAGCTCATCCAATTTTGACTTCTATTTGTAAAAAATGG 5339

Qy     181 tatatcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgccca 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5340 TATATCAATAGCTTTAAATATTTTAGAAATATGTACAAAGGGTTTTATTACAGCCGCCCA 5399

Qy     241 gataaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5400 GATAAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTTATTG 5459

Qy     301 ataaaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtacta 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5460 ATAAAAGAAAAGCCAGATTTAATATTATTAACGTTTCCTACACCAGTTATGTCGGTACTA 5519

Qy     361 actgagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacat 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5520 ACTGAGCAATTTAACATTAATATTCCAGTTGCTACAGTGATGACAGACTATCGCTTACAT 5579

Qy     421 aaaaactggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaa 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5580 AAAAACTGGATTACGCCGTATTCAACAAGATATTATGTGGCAACAAAAGAAACGAAACAA 5639

Qy     481 gacttcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgat 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5640 GACTTCATAGACGTAGGTATTGATCCTTCAACAGTTAAAGTGACAGGTATTCCTATTGAT 5699

Qy     541 aacaaatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatcca 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5700 AACAAATTTGAAACGCCTATTAATCAAAAGCAGTGGTTAATAGACAACAACTTAGATCCA 5759
```

```
Qy    601 gataagcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacg 660
Db   5760 GATAAGCAAACTATTTTAATGTCAGCTGGTGCATTTGGTGTATCTAAAGGTTTTGACACG 5819

Qy    661 atgattactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaag 720
Db   5820 ATGATTACTGATATATTAGCGAAAAGTGCAAATGCACAAGTAGTTATGATTTGTGGTAAG 5879

Qy    721 agcaaagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgatt 780
Db   5880 AGCAAAGAGCTAAAGCGTTCTTTAACAGCTAAGTTTAAATTAACGAGAATGTATTTGATT 5939

Qy    781 ctaggttataccaaacacatgaatgaatggatggcatcaagtcaacttatgattacgaaa 840
Db   5940 CTAGGTTATACCAAACACATGAATGAATGGATGGCATCAAGTCAACTTATGATTACGAAA 5999

Qy    841 cctggtggtatcacaataactgaaggtttcgcccgttgtattccaatgatttcctaaat 900
Db   6000 CCTGGTGGTATCACAATAACTGAAGGTTTCGCCCGTTGTATTCCAATGATTTCCTAAAT 6059

Qy    901 cctgcacctggtcaagagcttgaaaatgccttttactttgaagaaaaaggttttggtaaa 960
Db   6060 CCTGCACCTGGTCAAGAGCTTGAAAATGCCTTTTACTTTGAAGAAAAAGGTTTTGGTAAA 6119

Qy    961 a-cgctgatactccag 975
Db   6120 ATCGCTGATACTCCAG 6135
```

```
RESULT 3
AF270166
LOCUS       AF270166     3118 bp    DNA             BCT       01-AUG-2000
DEFINITION  Staphylococcus epidermidis strain SR1 clone step.1051e11 genomic
            sequence.
ACCESSION   AF270166
VERSION     AF270166.1  GI:9624074
KEYWORDS    .
SOURCE      Staphylococcus epidermidis.
  ORGANISM  Staphylococcus epidermidis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE   1  (bases 1 to 3118)
  AUTHORS   Kimmerly,W.J., Taylor,J.David, Nelsen,A.J., Godlevski,M.M.,
            Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
            Listenbee,S., Ashanti,C., Altshuller,G., Mamo,L., Shepherd,N.S.,
            Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
            Furdon,P.J.
  TITLE     Transposon-mediated sequencing of the Staphylococcus epidermidis
            genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3118)
  AUTHORS   Taylor,J.David, Kimmerly,W.J., Nelsen,A.J., Godlevski,M.M.,
            Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
            Listenbee,S., Ashanti,C., Altshuller,G., Mamo,L., Shepherd,N.S.,
            Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
            Furdon,P.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2000) Departments of Genomic Sciences and
            Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
            Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES             Location/Qualifiers
     source          1..3118
                     /organism="Staphylococcus epidermidis"
                     /strain="SR1"
                     /db_xref="taxon:1282"
                     /clone="step.1051e11"
BASE COUNT     1048 a    480 c    492 g    1098 t
ORIGIN

  Query Match             65.3%;  Score 636.2;  DB 2;  Length 3118;
  Best Local Similarity   78.9%;  Pred. No. 3.2e-104;
  Matches  758;  Conservative    0;  Mismatches  203;  Indels    0;  Gaps    0;

Qy      1 atggttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatg 60
Db    275 ATGGTTACTCAAAATAAAAAGATATTGATTATTACTGGATCATTTGGTAACGGTCATATG 334

Qy     61 caagttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtc 120
Db    335 CAAGTCACGCAAAGTATTGTCAACCAATTGAATGAGATGAATCTCAATCATTTATCAGTC 394

Qy    121 attgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatgg 180
Db    395 ATTCAACATGATTTGTTTATGGAAGCTCATCCAATTATGACTTCTATATGTAAGAAATGG 454

Qy    181 tatatcaatagctttaaatatttagaaatatgtacaaagggtttattacagccgccca 240
Db    455 TATATCAATAGCTTTAAATATTTAGAAATACATATAAACGATTTACTATAGTCGCCCT 514

Qy    241 gataaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattg 300
Db    515 AATGAGCTCGATAAATGTTTTTATAAATATTATGGATTAAATAAACTCATCAACTTACTT 574

Qy    301 ataaaagaaaagccagatttaatatattattaacgtttcctacaccagttatgtcggtacta 360
Db    575 ATTAAAGAAAAGCCTGATCTCATATTATTAACATTTCCAACACCTGTGATGTCAGTGTTG 634

Qy    361 actgagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacat 420
Db    635 ACCGAACAATTTAATATAAATATCCCTATTGCGACAGTTATGACAGATTATCGCATGCAT 694

Qy    421 aaaaactggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaa 480
Db    695 AAAAATTGGATTACACCATATTCACAAAGATATTATGTAGCAACAAAAGATACTAAAGAT 754

Qy    481 gacttcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgat 540
Db    755 GATTTCATTGAAGCTGGTGTTCCTGCTTCATATATTAAAGTGACGGGCATTCCTATTGCT 814

Qy    541 aacaaatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatcca 600
Db    815 GATAAATTTGAAGAATCTATTGATAAAGAAGAATGGTTATCGCAACAACATTTAGACCCT 874

Qy    601 gataagcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacg 660
Db    875 TCAAAACCTACTATATTAATGTCAGCAGGTGCATTTGGTGTTTCAAAAGGCTTTGACTAT 934

Qy    661 atgattactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaag 720
Db    935 ATGATTAATAATATTTTAGAAAAAAGTCCAAATTCGCAAGTGGTCATGATTTGTGGACGT 994

Qy    721 agcaaagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgatt 780
Db    995 AGTAAGGAACTTAAACGTTCATTAAAAGCTAAGTTCAAAGATAATCCAAGTGTAATAATA 1054

Qy    781 ctaggttataccaaacacatgaatgaatggatggcatcaagtcaacttatgattacgaaa 840
Db   1055 TTAGGATATACAAATCACATGAATGAGTGGATGGCATCAAGCCAACTAATGATTACAAAA 1114

Qy    841 cctggtggtatcacaataactgaaggtttcgcccgttgtattccaatgatttcctaaat 900
Db   1115 CCTGGTGGTATCACAATTTCCGAAGGACTTAGTCGTTGTATTCCTATGATTTTTTTAAAC 1174

Qy    901 cctgcacctggtcaagagcttgaaaatgccttttactttgaagaaaaaggttttggtaaa 960
Db   1175 CCTGCACCCGGTCAAGAACTTGAAAATGCATATTACTTTGAAAGTAAAGGATTTGGAAAA 1234

Qy    961 a 961
Db   1235 A 1235

RESULT 4
AF270394
```

```
LOCUS       AF270394     3493 bp    DNA             BCT       01-AUG-2000
DEFINITION  Staphylococcus epidermidis strain SR1 clone step.4046a02 genomic
            sequence.
ACCESSION   AF270394
VERSION     AF270394.1  GI:9624308
KEYWORDS    .
SOURCE      Staphylococcus epidermidis.
  ORGANISM  Staphylococcus epidermidis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE   1  (bases 1 to 3493)
  AUTHORS   Kimmerly,W.J., Taylor,J.David, Nelsen,A.J., Godlevski,M.M.,
            Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
            Listenbee,S., Ashanti,C., Altshuller,G., Mamo,L., Shepherd,N.S.,
            Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
            Furdon,P.J.
  TITLE     Transposon-mediated sequencing of the Staphylococcus epidermidis
            genome
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3493)
  AUTHORS   Taylor,J.David, Kimmerly,W.J., Nelsen,A.J., Godlevski,M.M.,
            Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
            Listenbee,S., Ashanti,C., Altshuller,G., Mamo,L., Shepherd,N.S.,
            Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
            Furdon,P.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2000) Departments of Genomic Sciences and
            Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
            Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES             Location/Qualifiers
     source          1. .3493
                     /organism="Staphylococcus epidermidis"
                     /strain="SR1"
                     /db_xref="taxon:1282"
                     /clone="step.4046a02"
BASE COUNT     1172 a    534 c    547 g   1240 t
ORIGIN
```

```
  Query Match             65.3%;  Score 636.2;  DB 2;  Length 3493;
  Best Local Similarity   78.9%;  Pred. No. 3.2e-104;
  Matches  758;  Conservative    0;  Mismatches  203;  Indels    0;  Gaps    0;

Qy      1 atggttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatg 60
          ||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||||||||
Db    384 ATGGTTACTCAAAATAAAAAGATATTGATTATTACTGGATCATTTGGTAACGGTCATATG 443

Qy     61 caagttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtc 120
          ||||| || || ||||| || || ||| | ||||| |||||||  | ||||||   |||
Db    444 CAAGTCACGCAAAGTATTGTCAACCAATTGAATGAGATGAATCTCAATCATTTATCAGTC 503

Qy    121 attgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatgg 180
          ||| | || ||||| ||||||||||||||||||||| |||||||||| ||||| ||||||
Db    504 ATTCAACATGATTTGTTTATGGAAGCTCATCCAATTATGACTTCTATATGTAAGAAATGG 563

Qy    181 tatatcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgccca 240
          ||||||||||||||||||||||||||||||  || ||| | ||||| || || ||||||
Db    564 TATATCAATAGCTTTAAATATTTTAGAAATACATATAAACGATTTTACTATAGTCGCCCT 623

Qy    241 gataaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattg 300
          || | || || |||||||||| ||||| |||||| | |||||  | || || ||| |
Db    624 AATGAGCTCGATAAATGTTTTTATAAATATTATGGATTAAATAAACTCATCAACTTACTT 683

Qy    301 ataaaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtacta 360
          || |||||||||| ||| | |||||||||| ||||| ||||| || ||||| ||  |
Db    684 ATTAAAGAAAAGCCTGATCTCATATTATTAACATTTCCAACACCTGTGATGTCAGTGTTG 743

Qy    361 actgagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacat 420
          || || |||||||| || ||||| ||  |||| ||||| |||||||| |||||| | |||
Db    744 ACCGAACAATTTAATATAAATATCCCTATTGCGACAGTTATGACAGATTATCGCATGCAT 803

Qy    421 aaaaactggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaa 480
          ||||| |||||||| || |||||| |||||||||||| |||||||||| || ||| |
Db    804 AAAAATTGGATTACACCATATTCACAAAGATATTATGTAGCAACAAAAGATACTAAAGAT 863

Qy    481 gacttcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgat 540
          || ||||| || |  ||| ||  | |||||    |||||||||| || |||||||||| |
Db    864 GATTTCATTGAAGCTGGTGTTCCTGCTTCATATATTAAAGTGACGGGCATTCCTATTGCT 923

Qy    541 aacaaatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatcca 600
          | |||||||||    ||||| || || |  | ||||||    |  |  | ||||| ||
Db    924 GATAAATTTGAAGAATCTATTGATAAAGAAGAATGGTTATCGCAACAACATTTAGACCCT 983

Qy    601 gataagcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacg 660
             || |  ||||| |||||||||||| |||||||||||||| || ||||| ||||||
Db    984 TCAAAACCTACTATATTAATGTCAGCAGGTGCATTTGGTGTTTCAAAAGGCTTTGACTAT 1043

Qy    661 atgattactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaag 720
          |||||||  | |||| |||| |||||| ||||| | ||||| || ||||||||||
Db   1044 ATGATTAATAATATTTTAGAAAAAAGTCCAAATTCGCAAGTGGTCATGATTTGTGGACGT 1103

Qy    721 agcaaagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgatt 780
          || || || || || ||||| |||| |||||||||| |||   |   ||      | ||
Db   1104 AGTAAGGAACTTAAACGTTCATTAAAAGCTAAGTTCAAAGATAATCCAAGTGTAATAATA 1163

Qy    781 ctaggttataccaaacacatgaatgaatggatggcatcaagtcaacttatgattacgaaa 840
          |||| ||||| ||.|||||||||| |||||||||||||| ||||| |||||||| |||
Db   1164 TTAGGATATACAAATCACATGAATGAGTGGATGGCATCAAGCCAACTAATGATTACAAAA 1223

Qy    841 cctggtggtatcacaataactgaaggtttcgcccgttgtattccaatgattttcctaaat 900
          |||||||||||||||||  | |||||  |    |||||||||| ||||||||  ||||
Db   1224 CCTGGTGGTATCACAATTTCCGAAGGACTTAGTCGTTGTATTCCTATGATTTTTTTAAAC 1283

Qy    901 cctgcacctggtcaagagcttgaaaatgccttttactttgaagaaaaaggttttggtaaa 960
          |||||||| |||||||| |||||||||| | ||||||||||   |||| |||| |||
Db   1284 CCTGCACCCGGTCAAGAACTTGAAAATGCATATTACTTTGAAAGTAAAGGATTTGGAAAA 1343

Qy    961 a 961
          |
Db   1344 A 1344
```

```
RESULT   5
AR106079/c
LOCUS       AR106079     2600 bp    DNA             PAT       14-FEB-2001
DEFINITION  Sequence 1 from patent US 6103507.
ACCESSION   AR106079
VERSION     AR106079.1  GI:12820144
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2600)
  AUTHORS   Shilling,L.Kathleen, Wallis,N.Gail and Zabierowski,S.
  TITLE     MurE
  JOURNAL   Patent: US 6103507-A 1 15-AUG-2000;
FEATURES             Location/Qualifiers
     source          1. .2600
                     /organism="unknown"
BASE COUNT      893 a    369 c    494 g    844 t
ORIGIN
```

```
  Query Match             44.8%;  Score 437;  DB 9;  Length 2600;
  Best Local Similarity   100.0%;  Pred. No. 1e-68;
  Matches  437;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 ATGGTTACTCAAAATAAAAAGATATTGATTATTACTGGCTCATTCGGTAACGGTCATATG 498

Qy     61 caagttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 CAAGTTACACAGAGTATCGTTAATCAACTTAATGATATGAATCTAGACCATTTAAGCGTC 438

Qy    121 attgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatgg 180
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 ATTGAGCACGATTTATTTATGGAAGCTCATCCAATTTTGACTTCTATTTGTAAAAAATGG 378

Qy    181 tatatcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgccca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 TATATCAATAGCTTTAAATATTTTAGAAATATGTACAAAGGGTTTTATTACAGCCGCCCA 318

Qy    241 gataaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 GATAAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTTATTG 258

Qy    301 ataaaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtacta 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 ATAAAAGAAAAGCCAGATTTAATATTATTAACGTTTCCTACACCAGTTATGTCGGTACTA 198

Qy    361 actgagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 ACTGAGCAATTTAACATTAATATTCCAGTTGCTACAGTGATGACAGACTATCGCTTACAT 138

Qy    421 aaaaactggattacgcc 437
          |||||||||||||||||
Db    137 AAAAACTGGATTACGCC 121


RESULT   6
E35634/c
LOCUS       E35634       2600 bp    DNA             PAT       07-FEB-2001
DEFINITION  MurE.
ACCESSION   E35634
VERSION     E35634.1  GI:13019110
KEYWORDS    JP 1999225779-A/1.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 2600)
  AUTHORS   Nicola,G.W.L.L. and Schilling,S.Z.
  TITLE     MurE
  JOURNAL   Patent: JP 1999225779-A 1 24-AUG-1999;
            SMITHKLINE BEECHAM CORP,SMITHKLINE BEECHAM CORP PUBLIC LTD CO
COMMENT     OS   Unidentified
            PN   JP 1999225779-A/1
            PD   24-AUG-1999
            PF   25-SEP-1998 JP 1998269356
            PR   25-SEP-1997 US    60/059956,10-DEC-1997 US    08/988251 PI
            NICOLA G WALLIS,LISA K SCHILLING,SUSAN  ZABIAROSUKI PC
            C12N15/09,A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00, PC
            A61K31/00,
            PC   A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K31/00,A61K35/74,
            PC   A61K38/00,
            PC   A61K39/085,A61K39/395,A61K39/395,A61K45/00,A61K48/00,
            C07K16/12, PC   C12N9/00,
            PC   C12Q1/68,G01N33/53//C12P21/08,(C12N15/09,C12R1:46),C12N15/00,
            PC   A61K37/02,
            PC   (C12N15/00,C12R1:46)
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1. .2600
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .2600
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      893 a    369 c    494 g    844 t
ORIGIN


  Query Match             44.8%;  Score 437;  DB 10;  Length 2600;
  Best Local Similarity   100.0%;  Pred. No. 1e-68;
  Matches  437;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 ATGGTTACTCAAAATAAAAAGATATTGATTATTACTGGCTCATTCGGTAACGGTCATATG 498

Qy     61 caagttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 CAAGTTACACAGAGTATCGTTAATCAACTTAATGATATGAATCTAGACCATTTAAGCGTC 438

Qy    121 attgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatgg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 ATTGAGCACGATTTATTTATGGAAGCTCATCCAATTTTGACTTCTATTTGTAAAAAATGG 378

Qy    181 tatatcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgccca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 TATATCAATAGCTTTAAATATTTTAGAAATATGTACAAAGGGTTTTATTACAGCCGCCCA 318

Qy    241 gataaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 GATAAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTTATTG 258

Qy    301 ataaaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtacta 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 ATAAAAGAAAAGCCAGATTTAATATTATTAACGTTTCCTACACCAGTTATGTCGGTACTA 198

Qy    361 actgagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 ACTGAGCAATTTAACATTAATATTCCAGTTGCTACAGTGATGACAGACTATCGCTTACAT 138

Qy    421 aaaaactggattacgcc 437
          |||||||||||||||||
Db    137 AAAAACTGGATTACGCC 121


RESULT   7
AX016296
LOCUS       AX016296     1149 bp    DNA             PAT       07-SEP-2000
DEFINITION  Sequence 1 from Patent WO9949052.
ACCESSION   AX016296
VERSION     AX016296.1  GI:10041859
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 1149)
  AUTHORS   Zaehringer,U., Heinz,E., Jorasch,P. and Wolter,F.P.
  TITLE     Processive glycosyltransferase
  JOURNAL   Patent: WO 9949052-A 1 30-SEP-1999;
            FORSCHUNGSZENTRUM BORSTEL (DE); ZAEHRINGER ULRICH (DE); HEINZ ERNST
            (DE); JORASCH PETRA (DE); WOLTER FRANK P (DE); GVS GES FUER ERWERB
            UND VERWER (DE)
FEATURES             Location/Qualifiers
     source          1. .1149
                     /organism="Bacillus subtilis"
                     /db_xref="taxon:1423"
BASE COUNT      380 a    217 c    244 g    308 t
ORIGIN


  Query Match             12.1%;  Score 118;  DB 9;  Length 1149;
  Best Local Similarity   48.5%;  Pred. No. 7.6e-12;
  Matches  458;  Conservative    0;  Mismatches  465;  Indels   21;  Gaps    4;

Qy     13 aataaaaagatattgattattactggctcattcggtaacggtcatatgcaagttacacag 72
          |||||||   |||| ||| | ||||     | ||| || || ||| |||| ||  |  |
Db     10 AATAAAAGAGTATTAATTTTGACTGCAAATTACGGAAATGGACATGTGCAGGTAGCCAAA 69

Qy     73 agtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattgagcacgat 132
          |   |  | | |||  |       |     | | ||| |||  ||   |      ||
Db     70 ACACTTTATGAACAATGTGTACGGCTCGGCTTTCAGCATGTAACAGTTTCT------AAT 123

Qy    133 ttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtatatcaatagc 192
          || |     ||  |  |||| ||| |  |     |   | || |    ||| | || |||
Db    124 TTGTACCAAGAGTCAAATCCGATTGTTTCAGAGGTAACTCAATACCTTTATTTAAAAAGC 183
```

```
Qy    193 tttaaatattttagaaatatgtacaaagggttttatt---------acagccgcccagat 243
          ||   |       | | |  | ||      ||||||||          |||    |    ||
Db    184 TTCTCAATCGGGAAACAGTTTTATCGTTTGTTTTATTACGGAGTTGACAAAATCTATAAT 243

Qy    244 aaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgata 303
          ||||   |      | |||||    |    ||   ||| |  ||     |  |  |
Db    244 AAACGTAAATTCAATATTTACTTTAAAATGGGTAATAAAAGATTGGGCGAACTTGTCGAT 303

Qy    304 aaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaact 363
           || |  |||| ||| | ||  |    || |||||  |     ||  || ||| |   |
Db    304 GAACATCAGCCCGATATTATTATTAATACATTTCCGATGATCGTCGTGCCGGAATACAGA 363

Qy    364 gagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataaa 423
            | |  |    |      |||||        |  || ||||| || | | |   | ||||||
Db    364 CGCCGAACTGGAAGAGTCATTCCTACCTTCAACGTTATGACTGATTTTTGTCTTCATAAA 423

Qy    424 aactggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagac 483
          |  ||| ||       |       | ||||||||||||| ||| |  |   |||  || |
Db    424 ATTTGGGTTCACGAAAACGTGGATAAATATTATGTGGCGACAGATTACGTGAAGGAAAAA 483

Qy    484 ttcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataac 543
           |  | ||  | || | | ||||    |  || ||| | ||||| ||||| ||
Db    484 CTGCTGGAGATCGGCACTCATCCAAGCAATGTAAAAATCACAGGAATTCCAATCAGGCCG 543

Qy    544 aaatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccagat 603
           ||||||||    | ||   |      | |   |  | | |  ||||   |    ||| |
Db    544 CAATTTGAAGAATCCATGCCTGTTGGCCCGATATATAAAAAGTACAATCTTTCACCAAAC 603

Qy    604 aagcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatg 663
          ||  ||    || | ||    || |||||    |||||||  ||      |  |   | ||
Db    604 AAAAAAGTGCTTCTGATCATGGCAGGTGCTCACGGTGTATTAAAGAACGTAAAAGAGCTG 663

Qy    664 attactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagc 723
            || |   | |  ||   ||   | |  |||||||| |  | || || || | |
Db    664 TGC---GAAAACCTTGTCAAGGATGACCAAGTGCAAGTAGTTGTCGTGTGCGGGAAAAAT 720

Qy    724 aaagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttga---tt 780
          |  |   ||||   |||||  |  ||          ||   |     | || |   ||
Db    721 ACGGCTTTAAAAGAATCTTTGAGTGCGCTTGAAGCGGAAAATGGTGACAAATTAAAAGTT 780

Qy    781 ctaggttataccaaacacatgaatgaatggatggcatcaagtcaacttatgattacgaaa 840
          || || |||    | | |||  ||||     |        |   |    |||||||| ||
Db    781 CTGGGCTATGTGGAGCGCATTGATGAGCTATTTCGGATCACAGATTGCATGATTACCAAG 840

Qy    841 cctggtggtatcacaataactgaaggtttcgcccgttgtattccaatgattttcctaaat 900
          || || || || ||  | || ||||     |||  | |  | ||  | ||| |    ||
Db    841 CCCGGCGGCATTACTTTGACAGAAGCCACAGCCATTGGAGTGCCTGTCATTCTGTACAAA 900

Qy    901 cctgcacctggtcaagagcttgaaaatgccttttactttgaaga 944
          || |  ||||| || ||    |||||||    | |||||||||
Db    901 CCCGTGCCTGGCCAGGAAAAAGAAAATGCAAACTTCTTTGAAGA 944
```

```
RESULT   8
BACYACA/c
LOCUS       BACYACA     35040 bp    DNA             BCT       10-APR-1996
DEFINITION  Bacillus subtilis (YAC10-9 clone) DNA region between the serA and
            kdg loci.
ACCESSION   L77246
VERSION     L77246.1  GI:1256615
KEYWORDS    DNA-binding protein; adenine phosphoribosyltransferase; bcsA gene;
            bsaA gene; cold shock protein; cspD gene; degR gene; dfrA gene;
            dihydrofolate reductase; dihydroxyacid dehydratase;
            exodeoxyribonuclease; homoserine O-succinyltransferase; ilvA gene;
            ilvD gene; metB gene; naringenin-chalcone synthase; stress-related
            protein; threonine dehydratase; thyB gene; thymidylate synthase;
            transport protein; xpt gene; ypaQ gene; ypbQ gene; ypbR gene; ypbS
            gene; ypcP gene; ypdP gene; ypdQ gene; ypeP gene; ypeQ gene; ypfP
            gene; ypgQ gene; ypgR gene; yphP gene; ypiP gene; ypjP gene; ypjQ
            gene; ypkP gene; yplP gene; yplQ gene; ypmP gene; ypmQ gene; ypmR
            gene; ypmS gene; ypmT gene; ypnP gene; ypoP gene; yppP gene; yppQ
            gene; ypqP gene; ypwA gene.
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 35040)
  AUTHORS   Iwakura,M., Kawata,M., Tsuda,K. and Tanaka,T.
  TITLE     Nucleotide sequence of the thymidylate synthase B and dihydrofolate
            reductase genes contained in one Bacillus subtilis operon
  JOURNAL   Gene 64 (1), 9-20 (1988)
  MEDLINE   88284366
REFERENCE   2  (bases 1 to 35040)
  AUTHORS   Sorokin,A.V., Azevedo,V., Zumstein,E., Galleron,N., Ehrlich,S.D.
            and Serror,P.
  TITLE     Sequence analysis of the Bacillus subtilis chromosome region
            between the serA and kdg loci cloned in a yeast artificial
            chromosome
  JOURNAL   Microbiology (1996) In press
REFERENCE   3  (bases 1 to 35040)
  AUTHORS   Capuano,V., Galleron,N., Pujic,P., Sorokin,A.V. and Ehrlich,S.D.
  TITLE     Organization of the Bacillus subtilis chromosome between kdg and
            the attachment site of the SPbetta prophage - use of Long Accurate
            PCR and yeast artificial chromosomes for sequencing
  JOURNAL   Unpublished (1996)
FEATURES             Location/Qualifiers
     source          1. .35040
                     /organism="Bacillus subtilis"
                     /sub_strain="168"
                     /strain="Marburg"
                     /db_xref="taxon:1423"
                     /clone="YAC10-9"
                     /clone_lib="P. Serror"
                     /map="205 degrees"
     gene            1. .1095
                     /gene="ypwA"
     CDS             <1. .1095
                     /gene="ypwA"
                     /note="putative"
                     /citation=[2]
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA96610.1"
                     /db_xref="GI:1256616"
                     /translation="EFNKRFITYWGYQEHPYDALLDLFEPGVTVKVLDQLFAELKEAI
                     IPLVKQVTASGNKPDTSFITKAFPKEKQKELSLYFLQELGYDFDGGRLDETVHPFATT
                     LNRGDVRVTTRYDEKDFRTAIFGTIHECGHAIYEQNIDEALSGTNLSDGASMGIHESQ
                     SLFYENFIGRNKHFWTPYYKKIQEASPVQFKDISLDDFVRAINESKPSFIRVEADELT
                     YPLHIIIRYEIEKAIFSNEVSVEDLPSLWNQKYQDYLGITPQTDAEGILQDVHWAGGD
                     FGYFPSYALGYMYAAQLKQKMLEDLPEFDALLERGEFHPIKQWLTEKVHIHGKRKKPL
                     DIIKDATGEELNVRYLIDYLSNKYSNLYLL"
     terminator      1167. .1199
                     /note="putative"
                     /citation=[3]
     gene            1426. .2010
                     /gene="xpt"
     CDS             1426. .2010
                     /gene="xpt"
                     /EC_number="2.4.2.7"
                     /function="purine biosynthesis"
                     /note="27% identity with E.coli adenine
                     phosphoribosyltransferase"
                     /citation=[3]
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="adenine phosphoribosyltransferase"
                     /protein_id="AAA96611.1"
                     /db_xref="GI:1256617"
                     /translation="MEALKRKIEEEGVVLSDQVLKVDSFLNHQIDPLLMQRIGDEFAS
                     RFAKDGITKIVTIESSGIAPAVMTGLKLGVPVVFARKHKSLTLTDNLLTASVYSFTKQ
                     TESQIAVSGTHLSDQDHVLIIDDFLANGQAAHGLVSIVKQAGASIAGIGIVIEKSFQP
                     GRDELVKLGYRVESLARIQSLEEGKVSFVQEVHS"
     gene            2007. .3323
                     /gene="ypaQ"
```

```
CDS             2007. .3323
                /gene="ypaQ"
                /function="pyrimidine salvage"
                /note="26% of identity to the Bacillus caldolyticus uracil
                transport prot. . .; putative"
                /citation=[3]
                /codon_start=1
                /transl_table=11
                /product="transport protein"
                /protein_id="AAA96612.1"
                /db_xref="GI:1256618"
                /translation="MRNGFGKTLSLGIQHVLAMYAGAIVVPLIVGKAMGLTVEQLTYL
                VSIDIFMCGVATLLQVWSNRFFGIGLPVVLGCTFTAVSPMIAIGSEYGVSTVYGSIIA
                SGILVILISFFFGKLVSFFPPVVTGSVVTIIGITLMPVAMNNMAGGEGSADFGDLSNL
                ALAFTVLSIIVLLYRFTKGFIKSVSILIGILIGTFIAYFMGKVQFDNVSDAAVVQMIQ
                PFYFGAPSFHAAPIITMSIVAIVSLVESTGVYFALGDLTNRRLTEIDLSKGYRAEGLA
                VLLGGIFNAFPYTAFSQNVGLVQLTGIKKNAVIVVTGVILMAFGLFPKIAAFTTIIPS
                AVLGGAMVAMFGMVIAYGIKMLSRIDFAKQENLLIVACSVGLGLGVTVVPDIFKQLPS
                ALTLLTTNGIVAGSFTAVVLNIVYNVFSKAKKIEQEADLAEQKTAV"
terminator      3322. .3341
                /note="putative"
                /citation=[3]
gene            3397. .4494
                /gene="bcsA"
CDS             3397. .4494
                /gene="bcsA"
                /EC_number="2.3.1.74"
                /note="25.6% of identity to the Lycopersicon esculentum
                (tomato) naringenin-chalcone synthase; alternate gene name
                ypbP; putative"
                /citation=[3]
                /codon_start=1
                /transl_table=11
                /product="naringenin-chalcone synthase"
                /protein_id="AAA96613.1"
                /db_xref="GI:1256619"
                /translation="MAFILSIGTSLPAYNVNQEKAAEFARYMFQHSFKDIDRLLSSFK
                NGQIHSRQFVKPIEWYKEGHSFEEKNQIYIEETLKHSRAAVRECLSHPEFFQEAIPYE
                KVEAVFFVSSTGLSTPSIEARLMNELPFSPYTKRIPIWGLGCAGGASGLARAAEYCKA
                YPEAFVLVISAELCSLTFQPEDKTKSNLIGTSLFGDGIAAALLCGEKADRRVSKLKLA
                PKIMDAQSVLMKQSEDVMGWDFTDQGFKVIFSRDIPTLVEKWLKTNVQIFLDKHKLSF
                HDISVFLAHPGGKKVIDAYIKSLGLSSEKLSSAQSILQKHGNMSSATILYVIKDHLQN
                GHKKEAERGLIGALGPGFSSELLLFSWEKGA"
gene            4498. .5004
                /gene="ypbQ"
CDS             4498. .5004
                /gene="ypbQ"
                /note="33.3% of identity in 60 aa with the Sacharomyces
                cerevisiae farnesyl cysteine carboxyl-methyltransferase;
                putative"
                /citation=[3]
                /codon_start=1
                /transl_table=11
                /protein_id="AAA96614.1"
                /db_xref="GI:1256620"
                /translation="MFWLLIAILIVQRAAEMAVARQNEQKVKKQGAIEFGESHYPYII
                TMHILFFLSLIAEVLLMNKQPSSWWLGIAAVILSVQIVRYWALCSLGAYWNTKILVVP
                GVELVKKGPYKWMKHPNYAVVILEILLIPLLYQAYVTMCLFSIFNAVLLSVRIRAEDK
                ALQEYSVK"
gene            5340. .8921
                /gene="ypbR"
CDS             5340. .8921
                /gene="ypbR"
                /note="26.7% of identity in 165 aa to a Thermophilic
                bacterium hypothetical protein 6; putative"
                /citation=[3]
                /codon_start=1
                /transl_table=11
                /protein_id="AAA96615.1"
                /db_xref="GI:1256621"
                /translation="MTDQNRKELLHKTGELYKQFIENQDEQRAAKLAAVMKKAADEEV
                YIAFTGHYSAGKSSLLNCLLMENILPTSPIPTSANLVVIRNGEKRVRLHTTDGACAEL
                EGTYQKDKVQQYCKDGEQIESVEIFDRYTEIDSGVAYIDTPGIDSTDDAHFLSAASIL
                HQADALFYVVHYNHVHAEENVKFLRSIKESIPNVYFIVNQIDRHDETETKFGDYQAQV
                EEMLCNEGISREALYFTSVTEPDHPFNQMGALREELSRIEQQSKSNMQALTEQKVRNL
                LKEHTEMLKKDETGAPSFAEQLNIHTGLVQSLRDQLDEAEKQMTEAEKRMQEEINRIL
                KNANLTPFEMRELAAAFLESQEPSFKTGFFFSKAKTAQERDKRRNAFFSDVAKRTEAE
                ADWHMIDTLHKLAKVFDVYTAESEKLIQAYRTPLDISIIEHAVKHGAAFSSEYVLQYT
                KDLAELIRKEAKREAADIIKVLSAMVKERVSKDVQTINDRLVQESEKLVFLQEQARLE
                NNAREKTDRLWAIWEEESACPMHIDTEWFKSKKTRVAAPEQKQGRSQLTAQPMPKSEI
                KMEQEMPLQDQIKRFYTLSDILGECSMLLKQTSAFRERVKRLEERKFTLALFGGFSSG
                KSSFANALVGERVLPSSPTPTTATINKITKPINGNLNKTANVVFKTEDDLTAEILQLT
                GIPKEPAGRSFTEKWEKAVKKNRLQEEHVKLISNFLLAYEKYQQYIQEQKKLTIPLSE
                LKPYVAEETTACAVKEVTVYYTCPLTEKGITIVDTPGASSMNKRHTELAFQYIKDADA
                FFYMTYYQHSFSKGDRSFLRKLGLVKESLSMDKMFFIINAADLAKDKTELETVTDYVS
                AELVKEGVYEPQLFTVSSKEELVGKPESFYNQFSKVRKHLDRFIEVDVKKASAAQLSS
                EADKLCETVFQLHQSQHQSREEKEAQKQCLMLSFERTAADIEKRRNSKTIIEKVKKDT
                REQLYHIAQRLSYFANDLLKSAFHPGLQNGDWKKNVSKAMTTALHEYLFEYIQEIKTL
                DVRMSGFIERHINEEWLDHFQKTLNEDGYFSVYAGDQHSNGIQLKEVEPEIEERAFEQ
                ELKEIKSPKQFFEQKGKATFIEAVRMKLTKITEAWIKNEEESLISHYTAHLRRLQEDM
                GEKAIAQITDQKETYLRGYAEGEHAKEIEMAYQACISWKNSDNTIKM"
gene            8986. .9243
                /gene="ypbS"
CDS             8986. .9243
                /gene="ypbS"
                /note="putative"
                /citation=[3]
                /codon_start=1
                /transl_table=11
                /protein_id="AAA96616.1"
                /db_xref="GI:1256622"
                /translation="MSEVHKAISAHSSKQHEHIKAFMRLENMRELAIEEAVAKCRNDE
                PYTTDAINEITEQMNQLAKKGIVPTRRLVSKEMVREYVSRM"
terminator      9474. .9505
                /note="putative"
                /citation=[3]
gene            9565. .10455
                /gene="ypcP"
CDS             9565. .10455
```

Query Match 12.1%; Score 118; DB 2; Length 35040;
Best Local Similarity 48.5%; Pred. No. 5.5e-12;
Matches 458; Conservative 0; Mismatches 465; Indels 21; Gaps 4;

```
Qy      13 aataaaaagatattgattattactggctcattcggtaacggtcatatgcaagttacacag 72
           ||||||| |||| ||| | |||| | ||| || || ||| |||| || | |
Db   14927 AATAAAAGAGTATTAATTTTGACTGCAAATTACGGAAATGGACATGTGCAGGTAGCCAAA 14868

Qy      73 agtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattgagcacgat 132
           |  |  | | |||  |      |     |  | ||| |||  ||   |       ||
Db   14867 ACACTTTATGAACAATGTGTACGGCTCGGCTTTCAGCATGTAACAGTTTCT------AAT 14814

Qy     133 ttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtatatcaatagc 192
           || |     ||  |  |||| ||| |  |     |   | || |    ||| | || |||
Db   14813 TTGTACCAAGAGTCAAATCCGATTGTTTCAGAGGTAACTCAATACCTTTATTTAAAAAGC 14754

Qy     193 tttaaatattttagaaatatgtacaaagggttttatt---------acagccgcccagat 243
           ||   |      | | |  | ||      ||||||||         |||   |    ||
Db   14753 TTCTCAATCGGGAAACAGTTTTATCGTTTGTTTTATTACGGAGTTGACAAAATCTATAAT 14694

Qy     244 aaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgata 303
           ||||  |      | |||||    |    ||   ||| |  ||     |  |  |
Db   14693 AAACGTAAATTCAATATTTACTTTAAAATGGGTAATAAAAGATTGGGCGAACTTGTCGAT 14634

Qy     304 aaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaact 363
           || |  |||| ||| | ||  |    || |||||  |     ||  || ||| |  |
Db   14633 GAACATCAGCCCGATATTATTATTAATACATTTCCGATGATCGTCGTGCCGGAATACAGA 14574

Qy     364 gagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataaa 423
             | |  |   |     |||||       |  || ||||| || | | |  ||||||
Db   14573 CGCCGAACTGGAAGAGTCATTCCTACCTTCAACGTTATGACTGATTTTTGTCTTCATAAA 14514

Qy     424 aactggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagac 483
           |  ||| ||       |       | |||||||||||| ||| |  |   |||  || |
Db   14513 ATTTGGGTTCACGAAAACGTGGATAAATATTATGTGGCGACAGATTACGTGAAGGAAAAA 14454
```

```
Qy     484 ttcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataac 543
           |  | ||  | || | | ||||    |  || ||| | ||||| ||||| ||
Db   14453 CTGCTGGAGATCGGCACTCATCCAAGCAATGTAAAAATCACAGGAATTCCAATCAGGCCG 14394

Qy     544 aaatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccagat 603
           ||||||||    | ||   |      | |   |  | | |  ||||  |    ||| |
Db   14393 CAATTTGAAGAATCCATGCCTGTTGGCCCGATATATAAAAAGTACAATCTTTCACCAAAC 14334

Qy     604 aagcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatg 663
           ||  ||    || | ||    || ||||||   |||||||  ||     |  |   | ||
Db   14333 AAAAAAGTGCTTCTGATCATGGCAGGTGCTCACGGTGTATTAAAGAACGTAAAAGAGCTG 14274

Qy     664 attactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagc 723
              || |   | |  ||   ||   | |  |||||||||  |  | || || || |
Db   14273 TGC---GAAAACCTTGTCAAGGATGACCAAGTGCAAGTAGTTGTCGTGTGCGGGAAAAAT 14217

Qy     724 aaagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttga---tt 780
           |  |   ||||    ||||| |  ||            ||   |     | || |   ||
Db   14216 ACGGCTTTAAAAGAATCTTTGAGTGCGCTTGAAGCGGAAAATGGTGACAAATTAAAAGTT 14157

Qy     781 ctaggttataccaaacacatgaatgaatggatggcatcaagtcaacttatgattacgaaa 840
           || || |||    | | |||  ||||     |      |   |    |||||||| ||
Db   14156 CTGGGCTATGTGGAGCGCATTGATGAGCTATTTCGGATCACAGATTGCATGATTACCAAG 14097

Qy     841 cctggtggtatcacaatAactgaaggtttcgcccgttgtattccaatgattttcctaaat 900
           || || || || ||  | || ||||     |||  | |  | ||  | ||| |    ||
Db   14096 CCCGGCGGCATTACTTTGACAGAAGCCACAGCCATTGGAGTGCCTGTCATTCTGTACAAA 14037

Qy     901 cctgcacctggtcaagagcttgaaaatgccttttactttgaaga 944
           || |  ||||| || ||    |||||||     | |||||||||
Db   14036 CCCGTGCCTGGCCAGGAAAAAGAAAATGCAAACTTCTTTGAAGA 13993


RESULT   9
BSUB0012
LOCUS       BSUB0012   213680 bp    DNA             BCT       26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 12 of 21): from 2195541
            to 2409220.
ACCESSION   Z99115 AL009126
VERSION     Z99115.1  GI:2634478
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 213680)
  AUTHORS   Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
            Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
            Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
            Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
            Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
            Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
            Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
            Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
            Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
            Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
            Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
            Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
            Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
            Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
            Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
            Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
            Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
            Danchin,A.
  TITLE     The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
  JOURNAL   Nature 390 (6657), 249-256 (1997)
  MEDLINE   98044033
REFERENCE   2  (bases 1 to 213680)
  AUTHORS   Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
            Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
            Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
            adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
            68 89 48
FEATURES             Location/Qualifiers
     source          1. .213680
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            complement(90. .782)
                     /gene="yoqL"
     CDS             complement(90. .782)
                     /gene="yoqL"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13977.1"
                     /db_xref="GI:2634479"
                     /db_xref="SPTREMBL:O34581"
                     /translation="MDSKKHKIINGYYHKNCISCKSWLPATEENFYSVKKNKDGLHSY
                     CKACVLKKAKESMLKNYDKQLERMRERNLLPGMKEAKKKYNSSLKKKKTQQIWQEKNK
                     LKLKNYRLQRDAHKKHNITDVQWQKCKDYFNNKCSYCGLKIEDHKILFKGTYIQSDFH
                     KEHVDHKGANDISNCIPACKSCNSSKHDFAFEEWYNSSNKNFSSERLLKIKEWLNRFK
                     EERQDSEWRTKG"
     gene            complement(822. .1025)
                     /gene="yoqK"
     CDS             complement(822. .1025)
                     /gene="yoqK"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13978.1"
                     /db_xref="GI:2634480"
                     /db_xref="SPTREMBL:O34326"
                     /translation="MKLIEERTYEERLYIYSDDIEARKHFVSELKPKGWMVDNCWVGI
                     DLNQIKQFYRFKRELTDDYVFNR"
     gene            complement(1045. .1560)
                     /gene="yoqJ"
     CDS             complement(1045. .1560)
                     /gene="yoqJ"
                     /function="unknown"
                     /note="similar to hypothetical proteins from B. subtilis"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13979.1"
                     /db_xref="GI:2634481"
                     /db_xref="SPTREMBL:O34359"
                     /translation="MKNPTMLKLKDKLLAVIEELITKENKYRFITGGALGTDQAACWC
                     VHILKKKHPHIKNIIATPFKEQDKVWSADQKMWYKRMLNVADEIINVEELDKYKVSGE
                     KPGEFSPAKMQKRNEYMIDHSEAIVAVYDGSKSGTRNCLNYAKKTYLGHQLWRLHPDF
                     NFELDITYFVG"
     gene            complement(1771. .1965)
                     /gene="yoqI"
     CDS             complement(1771. .1965)
                     /gene="yoqI"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB13980.1"
                     /db_xref="GI:2634482"
                     /db_xref="SPTREMBL:O34339"
                     /translation="MIKSIILPEENTKITVGKPINDESNTKVIAIYDYREEPEEAFWV
```

```
                HLSNGNDLFVDNHEVIVEYE"
gene            complement(2014. .2466)
                /gene="yoqH"
CDS             complement(2014. .2466)
                /gene="yoqH"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB13981.1"
                /db_xref="GI:2634483"
                /db_xref="SPTREMBL:O34999"
                /translation="MKRFILVLSFLSIIVAYPIQTNASPMPCSVILEPVDKNLKNAKG
                VALIYKVQLNPPSAARTNISILAVHLPAPSSFGNYDSYEGFATKPGEISWRFKLYPTP
                EEESPSWAGRIDTISAEMKNVKVQVRLSNSSTQKLGPSILTKNIESCY"
gene            complement(2549. .2806)
                /gene="yoqG"
CDS             complement(2549. .2806)
                /gene="yoqG"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB13982.1"
                /db_xref="GI:2634484"
                /db_xref="SPTREMBL:O35030"
                /translation="MEKMNLLKEITIFDLNKIKPGTKVQVTWYKGTEMEYTHNGEVII
                NNGEKFYYNYVDKEGYVGHCHVNALDLKNYPDSLIVEIKSK"
gene            complement(2851. .3054)
                /gene="yoqF"
CDS             complement(2851. .3054)
                /gene="yoqF"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB13983.1"
                /db_xref="GI:2634485"
                /db_xref="SPTREMBL:O34584"
                /translation="MKFAPMDKVKFKTASHLNKLRTLKKRVPELDDPLLGECWEFEED
                GLRQFDWEENYEFVARPKHFNWD"
gene            complement(3063. .3227)
                /gene="yoqE"
CDS             complement(3063. .3227)
                /gene="yoqE"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB13984.1"
                /db_xref="GI:2634486"
                /db_xref="SPTREMBL:O34429"
                /translation="MENNVLYGVYSTRSRKFCFGIEEPSKTKARKELFNRIGTDAYKW
                RFEIRKIKRK"
gene            complement(3281. .4036)
                /gene="yoqD"
CDS             complement(3281. .4036)
                /gene="yoqD"
                /function="unknown"
                /note="similar to phage-related DNA-binding protein
                anti-repressor"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB13985.1"
                /db_xref="GI:2634487"
                /db_xref="SPTREMBL:O34449"
                /translation="MESYLTVIEQNGQLLVDSREVAEMVGKRHTDLLRSIDGYVAILL
                NAKLRSVEFFLESTYKDATGRSLKHFHLTRKGCDMVANKMTGAKGVLFTAQYVSKFEE
                MEKALKARPSLIDTYLDMNEDERAIAYFTERKAKRELQEQLTLAEPKVEKYDRFLNTD
                GLMKIGQVAKAIGIKGMGQNNLFRFLRENKVLIDGTNKNAPYQKYVERGFFQVKTQET
                SVGIKTITLVTPKGADFIVDLLKKHGHKREIAS"
gene            complement(4077. .4484)
                /gene="yoqC"
CDS             complement(4077. .4484)
                /gene="yoqC"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB13986.1"
                /db_xref="GI:2634488"
                /db_xref="SPTREMBL:O34834"
                /translation="MEVNVKTNQREKFIRNGIPYDELDTQMIHLIDILNFKIGLKTRH
                CCFGHKPYEEIQVMFEDEVNIKEDQILELAELAGREWKGLQLSFSKWARFSPLMFNWS
                LVLSKRFRNPEDPNKYRYLRSVEEFFESYAAKK"
gene            complement(4491. .4829)
                /gene="yoqB"
CDS             complement(4491. .4829)
                /gene="yoqB"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB13987.1"
                /db_xref="GI:2634489"

  Query Match             12.1%;  Score 118;  DB 2;  Length 213680;
  Best Local Similarity   48.5%;  Pred. No. 4.6e-12;
  Matches  458;  Conservative    0;  Mismatches  465;  Indels   21;  Gaps    4;

Qy      13 aataaaaagatattgattattactggctcattcggtaacggtcatatgcaagttacacag 72
           ||||||| |||| ||| | ||||     | ||| || || ||| |||| || | |
Db 110225 AATAAAAGAGTATTAATTTTGACTGCAAATTACGGAAATGGACATGTGCAGGTAGCCAAA 110284

Qy      73 agtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattgagcacgat 132
           |  |  | | |||  |       |     | | ||| |||  ||   |       ||
Db 110285 ACACTTTATGAACAATGTGTACGGCTCGGCTTTCAGCATGTAACAGTTTCT------AAT 110338

Qy     133 ttatttatggaagctcatccaatttttgacttctatttgtaaaaaatggtatatcaatagc 192
           || |     ||  |  |||| ||| |  |     |   | || |    ||| | || |||
Db 110339 TTGTACCAAGAGTCAAATCCGATTGTTTCAGAGGTAACTCAATACCTTTATTTAAAAAGC 110398

Qy     193 tttaaatattttagaaatatgtacaaagggttttatt---------acagccgcccagat 243
           ||   |       | | |  | ||       ||||||||         |||    |    ||
Db 110399 TTCTCAATCGGGAAACAGTTTTATCGTTTGTTTTATTACGGAGTTGACAAAATCTATAAT 110458

Qy     244 aaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgata 303
           ||||   |       | |||||    |    ||   ||| |  ||      |  |  |
Db 110459 AAACGTAAATTCAATATTTACTTTAAAATGGGTAATAAAAGATTGGGCGAACTTGTCGAT 110518

Qy     304 aaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaact 363
           || |  |||| ||| | ||  |    || ||||| |     ||  || ||| |   |
Db 110519 GAACATCAGCCCGATATTATTATTAATACATTTCCGATGATCGTCGTGCCGGAATACAGA 110578

Qy     364 gagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataaa 423
           | |  |   |     |||||       |  || ||||| || | | |  | ||||||
Db 110579 CGCCGAACTGGAAGAGTCATTCCTACCTTCAACGTTATGACTGATTTTTGTCTTCATAAA 110638

Qy     424 aactggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagac 483
           |  ||| ||       |       | ||||||||||||| ||| |  |   |||  || |
Db 110639 ATTTGGGTTCACGAAAACGTGGATAAATATTATGTGGCGACAGATTACGTGAAGGAAAAA 110698

Qy     484 ttcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataac 543
           |  | ||  | || | | ||||     |  || ||| | ||||| ||||| ||
Db 110699 CTGCTGGAGATCGGCACTCATCCAAGCAATGTAAAAATCACAGGAATTCCAATCAGGCCG 110758

Qy     544 aaatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccagat 603
           ||||||||    | ||   |      | |  |  | | |  ||||  |    ||| |
Db 110759 CAATTTGAAGAATCCATGCCTGTTGGCCCGATATATAAAAAGTACAATCTTTCACCAAAC 110818

Qy     604 aagcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatg 663
           ||  ||     || | ||     || |||||     |||||||  ||      |  |  | ||
Db 110819 AAAAAAGTGCTTCTGATCATGGCAGGTGCTCACGGTGTATTAAAGAACGTAAAAGAGCTG 110878

Qy     664 attactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagc 723
                || |   | |  ||   ||   | |  ||||||||| |  | || || || |
Db 110879 TGC---GAAAACCTTGTCAAGGATGACCAAGTGCAAGTAGTTGTCGTGTGCGGGAAAAAT 110935

Qy     724 aaagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttga---tt 780
           |  |  ||||   ||||| |  ||           ||   |     | || |   ||
Db 110936 ACGGCTTTAAAAGAATCTTTGAGTGCGCTTGAAGCGGAAAATGGTGACAAATTAAAAGTT 110995
```

```
Qy    781 ctaggttataccaaacacatgaatgaatggatggcatcaagtcaacttatgattacgaaa 840
Db 110996 CTGGGCTATGTGGAGCGCATTGATGAGCTATTTCGGATCACAGATTGCATGATTACCAAG 111055

Qy    841 cctggtggtatcacaatcactgaaggtttcgcccgttgtattccaatgattttcctaaat 900
Db 111056 CCCGGCGGCATTACTTTGACAGAAGCCACAGCCATTGGAGTGCCTGTCATTCTGTACAAA 111115

Qy    901 cctgcacctggtcaagagcttgaaaatgccttttactttgaaga 944
Db 111116 CCCGTGCCTGGCCAGGAAAAAGAAAATGCAAACTTCTTTGAAGA 111159


RESULT  10
CEY53C12
LOCUS       CEY53C12   242893 bp    DNA             HTG       10-JUL-1998
DEFINITION  Caenorhabditis elegans chromosome II clone Y53C12, *** SEQUENCING
            IN PROGRESS ***, in unordered pieces.
ACCESSION   Z92859
VERSION     Z92859.1  GI:3242101
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 242893)
  AUTHORS   Lennard,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Jun 20, 1998 this sequence version replaced gi:1888376.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .242893
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="Y53C12"
BASE COUNT    79854 a  43516 c  41855 g  77668 t
ORIGIN


  Query Match             7.4%;  Score 72;  DB 83;  Length 242893;
  Best Local Similarity   44.7%;  Pred. No. 0.00071;
  Matches  321;  Conservative    0;  Mismatches  395;  Indels    2;  Gaps    1;

Qy     63 agttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcat 122
Db 112842 ATTTATTTAATTAATATTTATTTATGTAAATAATATTTATTTATTTAATTAATATTTATT 112901

Qy    123 tgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggta 182
Db 112902 TAATTAATATTTATTTATTTAATTAATATTTATTTATTTATTTAATTAATAATTATTTAT 112961

Qy    183 tatcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgcccaga 242
Db 112962 GTAAATAATATTTATTTATTTAATTAATATTTATTTAATTATTTAATTAATATTTATTTA 113021

Qy    243 taaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgat 302
Db 113022 TTTAATTAATATTTATTTATTTATTTAATTAATATTTATTTATTTAATTAATATTTATTT 113081

Qy    303 aaaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaac 362
Db 113082 ATGTAAATAATATTTATTTATTTAATTAATATTTATTTATTTATTTAATTAATATTTATT 113141

Qy    363 tgagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataa 422
Db 113142 TATTTAATTAATATTTATTTTTCTAATTAATATTTATTTATTTAATTAATAATATATAAT 113201

Qy    423 aaactggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaaga 482
Db 113202 AATAAATATTAATTAAATAAATAAATATTATTTAAATAAATAAATATTAATTAAATAAAT 113261

Qy    483 cttcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataa 542
Db 113262 AAATAAATATTAATTAAATAAATAAATATTAATTAAATATATAAATATT--AATTGAATA 113319

Qy    543 caaatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccaga 602
Db 113320 AATAAATAAATATTAATTAAATAAATAAATATTAATTAAATAAATAAATAAATATTAATT 113379

Qy    603 taagcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgat 662
Db 113380 AAATAAAAACAATTAATTAAATAAATAAATATTAATTAAATAAATAAATTTTAATTAAAT 113439

Qy    663 gattactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagag 722
Db 113440 AAATAAATAAATATTAATTAAATAAATAAATATTAATTAAATATATAAATATTAATTGAA 113499

Qy    723 caaagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgatt 780
Db 113500 TAAATAAATAAATATTAATTAAATAAATAAATATTAATTAAATAAATAAATATTAATT 113557


RESULT  11
CEY53C12D
LOCUS       CEY53C12D    1496 bp    DNA             INV       09-NOV-2000
DEFINITION  Caenorhabditis elegans cosmid Y53C12D, complete sequence.
ACCESSION   AL033123 Z92859
VERSION     AL033123.1  GI:3820757
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 1496)
  AUTHORS   none.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  REMARK    The C.elegans Sequencing Consortium.
            Erratum:[[published errata appear in Science 1999 Jan
            1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
            3;285(5433):1493]]
REFERENCE   2  (bases 1 to 1496)
  AUTHORS   Lennard,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is not the entire insert of clone Y53C12D.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
```

neighbouring submissions.
The start of this sequence (1. .117) overlaps with the end of sequence Z99277.
The end of this sequence (1390. .1496) overlaps with the start of sequence Z99278.
For a graphical representation of this sequence and its analysis see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=Y53C12D.

```
FEATURES             Location/Qualifiers
     source          1. .1496
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="Y53C12D"
BASE COUNT      607 a     92 c     94 g    703 t
ORIGIN
```

Query Match 6.2%; Score 60.2; DB 6; Length 1496;
Best Local Similarity 44.5%; Pred. No. 0.15;
Matches 281; Conservative 0; Mismatches 348; Indels 2; Gaps 1;

```
Qy      63 agttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcat 122
           | |||   |    ||  ||| | |  | ||| ||||  || ||    ||| |    | |
Db     846 ATTTATTTAATTAATATTTATTTATGTAAATAATATTTATTTATTTAATTAATATTTATT 905

Qy     123 tgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggta 182
           | |  |  ||||||||||  ||          ||||     | || ||   ||  ||
Db     906 TAATTAATATTTATTTATTTAATTAATATTTATTTATTTATTTAATTAATAATTATTTAT 965

Qy     183 tatcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgcccaga 242
             |  |  ||||  ||||| |  |||||  ||    |    ||||  |  |         |
Db     966 GTAAATAATATTTATTTATTTAATTAATATTTATTTAATTATTTAATTAATATTTATTTA 1025

Qy     243 taaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgat 302
           |  | |  | |  | ||| |  |  || |    | ||| | | |||||||||| | |   |
Db    1026 TTTAATTAATATTTATTTATTTATTTAATTAATATTTATTTATTTAATTAATATTTATTT 1085

Qy     303 aaaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaac 362
           |    || |        || || | |||||  |||  |      | |||  |       ||
Db    1086 ATGTAAATAATATTTATTTATTTAATTAATATTTATTTATTTATTTAATTAATATTTATT 1145

Qy     363 tgagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataa 422
           |    |||| |   ||| | ||| | ||  ||      |      | ||    |     |
Db    1146 TATTTAATTAATATTTATTTTTCTAATTAATATTTATTTATTTAATTAATAATATATAAT 1205

Qy     423 aaactggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaaga 482
           ||     ||||          | |   ||||       |  || |  ||  ||| ||
Db    1206 AATAAATATTAATTAAATAAATAAATATTATTTAAATAAATAAATATTAATTAAATAAAT 1265

Qy     483 cttcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataa 542
             | | |    |   | | |       |  | |||||   | |  ||||   |||||  |
Db    1266 AAATAAATATTAATTAAATAAATAAATATTAATTAAATATATAAATATT--AATTGAATA 1323

Qy     543 caaatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccaga 602
           | |  | ||        | ||| || |    |   | | | | |  ||   | ||  |
Db    1324 AATAAATAAATATTAATTAAATAAATAAATATTAATTAAATAAATAAATAAATATTAATT 1383

Qy     603 taagcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgat 662
           ||  |||  | || | |          |||   |  ||  | |  |||| |   ||
Db    1384 AAATAAAAACAATTAATTAAATAAATAAATATTAATTAAATAAATAAATTTTAATTAAAT 1443

Qy     663 gattactgatatattagcgaaaagtgcaaat 693
           | ||   | ||||||   |||     ||||
Db    1444 AAATAAATAAATATTAATTAAATAAATATAAT 1474
```

```
RESULT 12
ATMGDGSYN
LOCUS       ATMGDGSYN    1648 bp    mRNA            PLN       30-JUL-1998
DEFINITION  Arabidopsis thaliana mRNA for monogalactosyldiacylglycerol
            synthase.
ACCESSION   AJ000331
VERSION     AJ000331.1  GI:3367637
KEYWORDS    mgd gene; monogalactosyldiacylglycerol synthase.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1648)
  AUTHORS   Ohta,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUL-1997) Ohta H., Faculty of Bioscience and
            Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta
            Midori-ku Yokohama, Kanagawa/226, JAPAN
  REMARK    Revised by [3]
REFERENCE   2  (bases 1 to 1648)
  AUTHORS   Awai,K., Shimojima,M., Masuda,T., Takamiya,K.I. and Ohta,H.
  TITLE     cDNA cloning of an Arabidopsis monogalactosyldiacylglycerol
            synthase revealed that the enzyme belongs to an uridine
            diphosphate-utilizing glycosyltransferase superfamily
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1648)
  AUTHORS   Ohta,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1998) Ohta H., Faculty of Bioscience and
            Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta
            Midori-ku Yokohama, Kanagawa/226, JAPAN
COMMENT     On Aug 1, 1998 this sequence version replaced gi:3319653.
FEATURES             Location/Qualifiers
     source          1. .1648
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
     source          1. .1647
                     /organism="Arabidopsis thaliana"
                     /cultivar="columbia"
                     /db_xref="taxon:3702"
     gene            89. .1495
                     /gene="mgd"
     CDS             89. .1495
                     /gene="mgd"
                     /EC_number="2.4.1.46"
                     /function="glycosyltransferase"
                     /codon_start=1
                     /evidence=experimental
                     /product="monogalactosyldiacylglycerol synthase"
                     /protein_id="CAA04005.1"
                     /db_xref="GI:3367638"
                     /db_xref="SPTREMBL:O82730"
                     /translation="MATTVMALAEKVLERVYGTSKSAVSVTSGDGEKTHRHTHHHIHR
                     IKSYDDIDEDESSLELIQIGAERTKNVLILMSDTGGGHRASAEAIRDAFKIEFGDKYR
                     VIVKDVWKEYTGWPLNDMERSYKFMVKHVQLWKVAFHSTSPKWIHSCYLAAIAAYYAK
                     EVEAGLMEYKPEIIISVHPLMQHIPLWVLKWQELQKRVLFVTVITDLNTCHPTWFHPG
                     VNRCYCPSQEVAKRALFDGLDESQVRVFGLPVRPSFARAVLVKDDLRKELEMDQDLRA
                     VLLMGGGEGMGPVKETAKALEEFLYDKENRKPIGQMVVICGRNKKLASALEAIDWKIP
                     VKVRGFETQMEKWMGACDCIITKAGPGTIAESLIRSLPIILNDYIPGQEKGNVPYVVE
                     NGAGVFTRSPKETARIVGEWFSTKTDELEQTSDNARKLAQPEAVFDIVKDIDELSEQR
                     GPLASVSYNLTSSFASLV"
BASE COUNT      462 a    313 c    402 g    471 t
ORIGIN
```

Query Match 5.9%; Score 57.4; DB 13; Length 1648;
Best Local Similarity 51.1%; Pred. No. 0.46;
Matches 161; Conservative 0; Mismatches 151; Indels 3; Gaps 1;

```
Qy     642 atctaaaggttttgacacgatgattactgatatattagcgaaaagtgcaaatgcacaagt 701
           | |||||| | ||||     |  |   ||||| |      |  |   | | ||  ||| |
Db     964 AGCTAAAGCTCTTGAGGAATTTTTGTATGATAAAGAGAACAGGAAGCCTATTGGGCAAAT 1023

Qy     702 agttatgatttgtggtaagagcaaagagctaaagcgttctttaacagctaagtttaaatt 761
           ||| | || |||||     | |||  |  |      | | |||  ||| |   || | |
Db    1024 GGTTGTTATCTGTGGACGTAACAAGAAATTGGCATCTGCATTAGAAGCCA---TTGACTG 1080

Qy     762 aacgagaatgtatttgattctaggttataccaaacacatgaatgaatggatggcatcaag 821
```

```
Db   1081 GAAGATTCCTGTTAAGGTTCGAGGATTTGAGACTCAAATGGAGAAATGGATGGGAGCTTG 1140

Qy    822 tcaacttatgattacgaaacctggtggtatcacaataactgaaggtttcgcccgttgtat 881
          | |    || || || ||| ||||       |||||  |||||    |    || |   |
Db   1141 TGACTGCATCATCACAAAAGCTGGACCAGGAACAATCGCTGAATCGCTGATTCGATCACT 1200

Qy    882 tccaatgattttcctaaatcctgcacctggtcaagagcttgaaaatgccttttactttga 941
          ||| || ||  ||    ||      ||||| || |||   |  ||||     ||  | |
Db   1201 TCCTATCATCCTCAACGATTACATTCCTGGACAGGAGAAAGGGAATGTGCCGTATGTAGT 1260

Qy    942 agaaaaaggttttgg 956
            || || |||   ||
Db   1261 GGAGAATGGTGCAGG 1275
```

```
RESULT  13
CELH28G03
LOCUS       CELH28G03   44352 bp    DNA             INV       24-SEP-1999
DEFINITION  Caenorhabditis elegans cosmid H28G03.
ACCESSION   AF098501
VERSION     AF098501.2  GI:5921681
KEYWORDS    .
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 44352)
  AUTHORS   The C. elegans Genome Sequencing Consortium, Washington University
            Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,
            Hinxton, U.K.,C.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  REMARK    Erratum:[[published erratum appears in Science 1999 Jan
            1;283(5398):35]]
REFERENCE   2  (bases 1 to 44352)
  AUTHORS   Wamsley,P. and Gibson,A.
  TITLE     The sequence of C. elegans cosmid H28G03
  JOURNAL   Unpublished (1999)
REFERENCE   3  (bases 1 to 44352)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 44352)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 24, 1999 this sequence version replaced gi:3786414.
            Submitted by:
                    Genome Sequencing Center
                    Department of Genetics, Washington University,
                    St. Louis, MO 63110, USA, and
                    Sanger Centre, Hinxton Hall
                    Cambridge CB10 IRQ, England
                    e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


            NOTICE:  This sequence may not be the entire insert of this clone.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we provide a small overlap between
            neighboring submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
             chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems,
            such as compressions and repeats; all regions were covered by
            sequence from more than one m13 subclone.
```

NEIGHBORING COSMID INFORMATION:

The 5' clone is T03G11, 4800 bp overlap;3' clone is Y34B4A, 6600 bp overlap. Actual start of this clone is at base position 78 of CELH28G03; actual end is at 37952 of CELH28G03

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation).

```
FEATURES             Location/Qualifiers
     source          1. .44352
                     /organism="Caenorhabditis elegans"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /clone="H28G03"
                     /chromosome="X"
     gene            6742. .16476
                     /gene="H28G03.2"
     CDS             join(6742. .7098,7438. .7536,7588. .7643,12371. .12518,
                     15241. .15466,15654. .15705,15761. .16121,16167. .16298,
                     16354. .16476)
                     /gene="H28G03.2"
                     /note="coded for by C. elegans cDNA yk143c10.5; coded for
                     by C. elegans cDNA yk143c10.3; coded for by C. elegans
                     cDNA yk196a6.3; coded for by C. elegans cDNA yk196a6.5;
                     coded for by C. elegans cDNA yk255d2.3; coded for by C.
                     elegans cDNA yk255d2.5; coded for by C. elegans cDNA
                     yk413e1.5; coded for by C. elegans cDNA yk447a9.3; coded
                     for by C. elegans cDNA yk447a9.5; coded for by C. elegans
                     cDNA yk449a11.3; coded for by C. elegans cDNA yk449a11.5;
                     coded for by C. elegans cDNA yk467c2.3; coded for by C.
                     elegans cDNA yk467c2.5; coded for by C. elegans cDNA
                     yk470g12.3; coded for by C. elegans cDNA yk470g12.5; coded
                     for by C. elegans cDNA yk480a11.3; coded for by C. elegans
                     cDNA yk480a11.5; coded for by C. elegans cDNA yk483e7.3;
                     coded for by C. elegans cDNA yk483e7.5; coded for by C.
                     elegans cDNA yk492a8.3; coded for by C. elegans cDNA
                     yk492a8.5; coded for by C. elegans cDNA yk494b5.3"
                     /codon_start=1
                     /protein_id="AAC67404.1"
                     /db_xref="GI:3786418"
                     /translation="MQRDNVIDSAGNPPVENPDIDLQDGEIGEESFEATQAYLQDMEN
                     SAEAIALERQRAEEAVMATMDEDDDGLNEFGRPRNPPRTPPEPAPEGQPTCESIGFVN
                     EEFDDVSDDGEIPGRKRIGPKTPPEEPESPKQIEQTREQGSDGELTGDDMEPIPIIEP
                     KLNKESKREKRHATREVHSRKRRHDDKKRSTVRPEVSKELTFPVVATQKCLSLLEFVE
                     KRGLTITDDELLQVVRDSIELDQAMGKIVKKKEAMKKQMEDMAAWELEKIKRIHANMP
                     RHMQDVLKFDGKTVHIAQERPPMPPMPFSASSLVSLKYGQPMPHHMGMPPPGMGPPPF
                     MPPPIGMPPPPLGMPPPHIGLGAAPYAVPPPMSLLGPPPFSVPPSVPPPTSSAAAPIV
                     PPPPVQSTAQPPPSGTSEAEKMKMTHGLNPIAFSNAPPGLKLGMNDYNQPPPAKKPAA
                     QQNPQMNRITNNLSSMLTNALKAQVSKAQNSLNNSSPTSTAPKKPVPSLMSINIPGVP
                     KAGSSGSQN"
     gene            17763. .18923
                     /gene="H28G03.1"
     CDS             join(17763. .18017,18067. .18228,18281. .18781,18831. .18923)
                     /gene="H28G03.1"
                     /note="Contains similarity to Pfam domain: PF00076 (rrm),
                     Score=105.1, E-value=4.4e-28, N=2; coded for by C. elegans
                     cDNA yk171a8.5; coded for by C. elegans cDNA yk171a8.3;
                     coded for by C. elegans cDNA yk295h11.5"
                     /codon_start=1
                     /protein_id="AAC67401.1"
                     /db_xref="GI:3786415"
                     /translation="MNHECLILRRMHISDLITILELPYLWDTFKVSKKNNLKLGVRSI
                     FKYALFQQFQQMGMGQYQHDSPPQLRKLFIGGLSHDTTDEQLGNYFSQWGPVVDAIVI
                     RDPNTKHSRGFGFVTFASIFSAESAMNDRPHKLGGKTVDSKRAIPREQMSSMIPPPFF
                     ETDPAPGCKLLLNGITNGVHSVDSLRVYFETFGTLDQVEILGQPRGLGFVIYEDKESA
                     DRCLAHNSGRHIVNERKIEVRVFTKHPNGSTYWKRPQSQSHSQRDLFEQLSQLNLKDG
                     DRKSTGNSSSAADTPQNFDEDSNYGGTTTEDDCNVFEHEEGSSEESSTEQTLENEKEN
                     SD"
     gene            complement(21324. .22960)
```

```
                /gene="H28G03.3"
CDS             complement(join(21324. .21442,21532. .21738,21823. .21870,
                22258. .22314,22364. .22439,22516. .22554,22598. .22636,
                22946. .22960))
                /gene="H28G03.3"
                /note="contains similarity to intermediate filament
                proteins (Pfam: filament.hmm, score: 15.43)"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC67402.1"
                /db_xref="GI:3786416"
                /translation="MLHHQAQRFNLKIENLKTMIPLHVHHLHYTVKLIFTEYSPHSPT
                YNPDSPSYGEPSVAAEQQRPASSARYNVGPANNFAVQTPSPAQIAAWGYSGVSTVEYE
                QMKNKNEYLQREHAEMQKSIKRKDREIERLRSKNRRLEEEGRNIREQMPSTRTCQSRS
                IYTWIFQIIMFLQHIFHESNNMSQKFKFSQNHLIITIFF"
gene            complement(23327. .25416)
                /gene="H28G03.4"
CDS             complement(join(23327. .23464,23548. .23591,23709. .24369,
                24510. .24685,24923. .25036,25170. .25416))
                /gene="H28G03.4"
                /note="Contains similarity to Pfam domain: PF00292 (PAX),
                Score=10.2, E-value=0.028, N=1"
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAC67403.1"
                /db_xref="GI:3786417"
                /translation="MNTSHHVSDEPVKLFKTSTSKNHILDEPFIIRLHLIDPSIHHIH
                RGKFRLFENDFKNEKYRCAEIIGNLFARANLRFFRWLSHGVSARNISRRPNVRFHLVY
                KTITQFAELGLLVDRFGRCMRASPMREPIVRFHRNGVAAKSIARRLKVSEKLVSTTIA
                RFKELGNFSDRSGRGKPPTAYKKSTCQFLSEAAKIKRKDRAMNLLRRFRNGAHRKVLF
                TDEKIFCIEQSFNTQNDRVYAKTQPNSRVQRTGYPKGIMVFAGITANGKTPLIFVPQG
                IKVNGNNYLDMLKTELMPWVKKHFKKTKWTFQQDGAPAHKHKNVQAWCESNFPDFIAF
                NQWPPSSPDLNPMDYSVWSVLEAKACSKPHRNIDSLKDSLKKAWDELDINYLRATVDS
                FPRRLEASRELLTIILTHPVWSKVWNELEIPYLRATVDAFPKRVRVCIEADGDIFELM
                EFCLFFSGE"
gene            complement(35875. .44307)
                /gene="H28G03.6"
CDS             complement(join(35875. .35970,36016. .36302,36349. .36498,
                37005. .37060,37112. .37340,37388. .37724,37774. .38040,
                38522. .38760,38806. .38926,38970. .39421,39473. .39782,
                40206. .40306,40368. .40926,40975. .41121,41199. .42050,
                42271. .42471,42529. .42622,42673. .42805,42990. .43284,
                43778. .43968,44229. .44307))
                /gene="H28G03.6"
                /note="contains similarity to Pfam domain PF00169 (PH),
                Score=19.5, E-value=4e-05, N=1; coded for by C. elegans
                cDNA yk53a7.3; coded for by C. elegans cDNA yk47d2.3;
                coded for by C. elegans cDNA yk126b8.3; coded for by C.
                elegans cDNA yk9f2.3; coded for by C. elegans cDNA
                yk194d7.3; coded for by C. elegans cDNA yk194d7.5; coded
                for by C. elegans cDNA yk421c6.3"
                /codon_start=1
                /protein_id="AAC67405.2"
                /db_xref="GI:5921682"
                /translation="MRDPDKVKSGPICDTVAVIVLEESDDENALPDVLHEVQSPHTSD
                NIPTSSIKKFARPRGWYNQSVSSPSEFFYQILTTERGTRRIAYVLSTWEEDEKTLNFK
                AVSIVLISQNFHPKAFKEILLEISNDLRTPEFSSSSELIRFLTYELVEEGSTIEIRTK
                TLHVELGFELIPISPVTGKDVAMLFKMLGFQNVIKIIHALLSDCRIVLASSSLMRLSR
                CQNAILSLLYPFEYVHSCVTILPDSLAEVLESPTPFLIGVLSEFVTSFGDENIVVYLD
                NGEVHVPDHAEIYKSDDYYYNSLHQRLRDVMFTTTSQEDLSIPNEERIEVDDFILDKK
                LRACFIYYFAELLYGYQYYILYTRIKGNFEKKLTTSLTFHVGAFRGFRKLTDMMSSSL
                LKSVYFQTFILTRALPRRKHDLFDEISCFKELDQLIFKQNSTSSESKKIIEHISCELI
                QKERYMEKCSARKQEIFTKIHWISGKELAQNNNSIIHTVKPKMRSNVILQAMLPVVNT
                HAEYHANQFEAYAHRIEALRNCLAAIFEGKVAFASKSLDAVKSSMRFAPLRIELCRLL
                NQKCSHDKLTDKQFEDIALLMNAALQAECEEDKDGVVRSLMYLSNVYSRKVAQGMQQY
                MYTAVQEHKVWKNQRFWTSCFYYEVHEMLFSEMLQKDRKITESLWCHTLRPCAMEMIN
                TDDTDQEELVKQENEMIQAQAKHFANILISLQIPLSEEFFEHEDAHRSVLNEKCKWIV
                NTLDSILGVTGRINGLSLSRIQTYVEAHVESLRDVYVEMSTGEHLKKGNFDPVLAHGE
                FLISDPIDCYLLTSIEESEMSLNRLENLLPADGSLFLTNYRVIFKGKSVDINATNGTI
                VQTIPLYSMESFKKLTNKKLIPTQLIEKGVKIEHIISIRSSCASSIIIAFDEDEINNM
                AIEKFLEVIETNSHNSFAFYNTRKDMKVVENGSHKFGTLNSAIRGFTKKKTDTRRIRS
                HSSHRGSIQLSFDKMEELDYLKKNAHIRYAVIDYPRIGLNSKIVKLRMSHSNLDYTIC
```

```
  Query Match             5.8%;  Score 56.8;  DB 6;  Length 44352;
  Best Local Similarity   45.6%;  Pred. No. 0.43;
  Matches  199;  Conservative    0;  Mismatches  237;  Indels    0;  Gaps    0;

Qy      56 atatgcaagttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaa 115
           | ||   | ||| | | |  |    |||| || | ||| |     |  || |  | | ||
Db   32632 AAATATTAATTAAATAAATAAATAATAATTAAATAAATAAATAATAATTAAATAAATAAA 32691

Qy     116 gcgtcattgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaa 175
           | ||| |  |  |  |||| ||  ||        |  |     |  | |    |||
Db   32692 TATTAATTAAATAAAAAATATTAATTAAATAAATAAATAAATATTAATTAAATAAATAAA 32751

Qy     176 aatggtatatcaatagctttaaatattttagaaatatgtacaaagggttttattacagcc 235
           |     ||   | |    |||||||| |  |||||  || | |    || | || |
Db   32752 TAAATATTAATTAAATAAATAAATATTATTTAAATAAATAAATATTAATTAAATAAATAA 32811

Qy     236 gcccagataaactagacaaatgtttttacaaatactatggacttaataagttaattaatt 295
             || || || | |||| ||  |  ||| | || | | | | ||| | ||| |||
Db   32812 ATAATAATTAAATAAATAAATATTAATTAAAAAAATAAGTATTAATTAAATAAATAAATA 32871

Qy     296 tattgataaaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcgg 355
             || | | ||    | ||||    || ||  ||   || | | | |
Db   32872 ATAATTAAATAAATAAATATTAATTAAATAAATAAATATTAATTAAATTAATAAATAAAT 32931

Qy     356 tactaactgagcaatttaacattaatattccagttgctacagtgatgacagactatcgct 415
           |||  |  |||  |   | ||||||  | |   || |      | | |  |
Db   32932 ATTAAACAAATAAATAAATATTAAATATTTAAATAAATAAATATTAAATAAATAAATAAA 32991

Qy     416 tacataaaaactggattacgccgtattcaacaagatattatgtggcaacaaaagaaacga 475
           ||   |  || |  || |     |||| | |||   | ||  |    |  ||  |||  |
Db   32992 TATTAATTAATTAAATAAATAAATATTAATCAAATAAATAAATAAATATTAATTAAATAA 33051

Qy     476 aacaagacttcataga 491
           |  || | |   || |
Db   33052 ATAAATATTAATTAAA 33067
```

```
RESULT  14
CEY18D10A
LOCUS       CEY18D10A  152878 bp    DNA             INV       09-NOV-2000
DEFINITION  Caenorhabditis elegans cosmid Y18D10A, complete sequence.
ACCESSION   AL034393 AL008871
VERSION     AL034393.1  GI:3979928
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 152878)
  AUTHORS   none.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  REMARK    The C.elegans Sequencing Consortium.
            Erratum:[[published errata appear in Science 1999 Jan
            1;283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
            3;285(5433):1493]]
REFERENCE   2  (bases 1 to 152878)
  AUTHORS   Harris,B.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1998) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, England and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
```

```
          IMPORTANT: This sequence is not the entire insert of clone Y18D10A.
          It may be shorter because we only sequence overlapping sections
          once, or longer because we arrange for a small overlap between
          neighbouring submissions.
          The true left end of clone F08A8 is at 152766 in this sequence. The
          true right end of clone B0019 is at 105 in this sequence. The start
          of this sequence (1. .105) overlaps with the end of sequence
          AL008866.
          The end of this sequence (152766. .152878) overlaps with the start
          of sequence Z99710.
          For a graphical representation of this sequence and its analysis
          see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
          name=Y18D10A.
FEATURES            Location/Qualifiers
     source         1. .152878
                    /organism="Caenorhabditis elegans"
                    /db_xref="taxon:6239"
                    /chromosome="I"
                    /clone="Y18D10A"
     gene           complement(join(7920. .8081,9752. .9924,10879. .11164,
                    12369. .12699,13906. .14204,15129. .16655,17748. .18104,
                    19860. .19949,20008. .20088,20709. .20807,22205. .23127,
                    24486. .24744,25724. .25909,26020. .26130,26236. .26256))
                    /gene="Y18D10A.1"
     CDS            complement(join(7920. .8081,9752. .9924,10879. .11164,
                    12369. .12699,13906. .14204,15129. .16655,17748. .18104,
                    19860. .19949,20008. .20088,20709. .20807,22205. .23127,
                    24486. .24744,25724. .25909,26020. .26130,26236. .26256))
                    /gene="Y18D10A.1"
                    /note="predicted using Genefinder
                    cDNA EST yk541b5.5 comes from this gene"
                    /codon_start=1
                    /protein_id="CAA22308.1"
                    /db_xref="GI:3979929"
                    /db_xref="SPTREMBL:Q9XW25"
                    /translation="MDEDIAEAQCASEGCANVRKNLGLLLMQNKQRAAEMQDFSKVQA
                    ECEVLRMKLVSTDAAYRESVQREAEAINAKEIAEASLKQGQYDSEFYKKKSASLQEQA
                    DDAEKHKDQANYWRNKYDSVCALTERLEANNIDIKSKYESTVKLLETAGKRMTSLDNQ
                    VTTLKAAKIENSSKTKKLDKYLKSVIQVVDALSSTSNLSKQSKSIKDLVTVFRRDDVR
                    LFFNSQPWKNSLSTSDPSISATEAEASEDPEHLDDVITEPAPPIGEQTLSDDEEEEEEE
                    VPEDEADDEVDREIRQLENELLQRVKEKTVRRPDFTAKMIELKAKSRAPRLVETKFVK
                    KWPEGIHKPTEKDSFGLLNSTKIHRNQFPTSDFETIAQATAERKKALLGAQGAAGASE
                    PGSSSSIHGKKGKTVRELQTEMMKGESVKEKAARMRAEASAGRSQAPGPAPAASEL
                    QDPPQDFGLSMSDPGSDSETEEHQKQEKHIPAMVTRRSARLSALPVTPKKASSSSKMP
                    PPPSPSPSTPGRRGRRPRTLSTMSMEPAAAAVTPAPRGRPRSRSAAKVSENTEPLSEA
                    PSAPVKRGRGRPRSRSTMSITEDSEPSTSSTAAKRSKRAESDEEEEQDLKLTNKSPEK
                    PKKPSKTTEETVGDVLKKRLRDTAKTTATVIHTPGPPLRTRKMERMRAPTAVTSSKKE
                    KPKNAGSADSSINEEEHEDETMILEEQTLDLPQQTSQQEPRISCGSELLDEQFDASEE
                    HSGTVPSAPELTKNPAPPVPEASEASAEPPKIDIPEQATPILALALALPTVSPTALEP
                    PKAQENPTAELPTTSEISGRAPQALPTSSQTPPTSGSAAPPVDDLLSEILSGAKTTKT
                    RKAAPPAVQKSISSTTQQAPPTSVQAPPTSCSAAPPVDDLLSEILSGAKTTKTTKTTK
                    MPPVDQKKISSEAPPISDSAPTSVHQQTPKSPKQILNSKYGLDISDSEDEEEEEEERG
                    MEIVEEEEEAPPISDSLQASEPSSTATVKPEKVVAVVKIFSPEIDSTSVEAPPEASVP
                    PKAPAATKIDNQLADQQASEPEPPKARKLPIARKIPPKIKISLPAPSSSTTSDDDLMS
                    DILAGAKTTKTTKPKAPPTRVAQTTRTKNLAQKRKASPPTPAGTTAPKRQYIKKSIDS
                    VLPPSSSSSTEPPSAPDSASTTSSMKKGGGAIMIEAVPCRPGGKAIKREQKPIGMKEM
                    MVQNVEKGGKKVNKIKTHLRQALDLKIPFEELKRPMEEKGIKLGDSIPLTPSDAVDVM
                    MEFLRETSAADMWAVLNRQRIDANLKPLMNKEEENFLQVSVTLNEHDQQLLEVFISRI
                    IFELCARESIGSNECGRLIRLFCHAVRFSESAENPDDVTSPVYQKKSTWIRRLFQILL
                    LNHPNQIGKSTCYILMSDVAKYCRFLVVEEMDKDHMGSEFHLAFRVLMHKDSEQASAV
                    NWLLNAKFDVAYVAPPTSNDVTEACINAHKQCIDDDDVTPSLLKKSIFLAKTAAPEVL
                    NVAFDFLRQKMDLIEEQFLKPKEMEKEPKRITAFSVPTVVTKMTKSEAEMLKKESKWQ
                    KNVLRVMIDNPFSSHSHAIFAALKQISPQVSQLREIIESSEINQYVEMTGEDAIRDAV
                    QNLVDLISIYTNFPLKPTTLPPSS"
     gene           complement(join(26502. .26573,27346. .27405,27490. .27580,
                    27624. .27788,28732. .29006))
                    /gene="Y18D10A.2"
     CDS            complement(join(26502. .26573,27346. .27405,27490. .27580,
                    27624. .27788,28732. .29006))
                    /gene="Y18D10A.2"
                    /note="predicted using Genefinder"
                    /codon_start=1
                    /protein_id="CAA22309.1"
                    /db_xref="GI:3979930"
                    /db_xref="SPTREMBL:Q9XW24"
                    /translation="MQMNPNKSPPISINFPLFTMSKFLLFISISAVATASDLVEVFGI
                    SRCPDTSKFIHNQLVPFYQNYKGNLSDGLKLDFHAVPTGGHQVDGKYVNRCLHGALEC
                    ALNKLQMCSKKHIKQDWLVTAGCIQGKTAYSAGLKCLPDTEEGKIVQNCAESEEGEYL
                    LNDENSYRYNVAPHSAWLPWIQVNGERNRNAEFKLKDFSSVFRLFFEVFLQKCLKRSI
                    KN"
     gene           complement(join(29879. .30346,31220. .31296,31346. .31522,
                    34557. .34852,35323. .35351))
                    /gene="Y18D10A.3"
     CDS            complement(join(29879. .30346,31220. .31296,31346. .31522,
                    34557. .34852,35323. .35351))
                    /gene="Y18D10A.3"
                    /codon_start=1
                    /protein_id="CAA22319.1"
                    /db_xref="GI:3979940"
                    /db_xref="SPTREMBL:Q9XW15"
                    /translation="MDGSGEKAIGFLIQFSLNARFWTDKQLLATDEPFSSENYQMVRR
                    FRIISEKRPKSVDKLYKTGEIPARGRGNKVGGRVFSKMPISKSMELSLVKLKQTPIIQ
                    QFYMKNGKKSTMVHIISKKTVSFIGQKLAAQIDEQLFTKYGFKVEQLMELAGLAAAQA
                    IAAHYPKSNVAVLCGPGNNGGDGFVCARHLQQFGFTPSIVYPKESRNELMKSLVVQCE
                    TSSIPITATLPTNLQAFPLIVDALFGFSFHPPTREPFTEMLKTVRASGIHVFSIDVPS
                    GWDVELGAPSGNDDDVIHPHSVISLTLPKLCMKNWTGPHFLGGRFVPKSLVDEHELLM
                    PQYPGFEQIVKLED"
     gene           join(35614. .35976,36025. .36138)
                    /gene="Y18D10A.4"
     CDS            join(35614. .35976,36025. .36138)
                    /gene="Y18D10A.4"
                    /note="predicted using Genefinder"
                    /codon_start=1
                    /protein_id="CAA22310.1"
                    /db_xref="GI:3979931"
                    /db_xref="SPTREMBL:Q9XW23"
                    /translation="MTGERQTEAVSKRSNVVSERCGGPPCLRLVDGRTLSVCVSDDSS
                    HIDGSIGYHFFSLDCLSLSLTLSTAAVVATATSSSSSPKAFGVSVSLFLWSLLVAWYV
                    VGVLLVTGGMDREKIYVWQFFHRLPTKRKIENSFAFSSFFFCFLKPMIIGFWTLSF"
     gene           join(39072. .39164,40245. .40513,41360. .41936,44226. .44375)
                    /gene="Y18D10A.5"
     CDS            join(39072. .39164,40245. .40513,41360. .41936,44226. .44375)
                    /gene="Y18D10A.5"
                    /note="predicted using Genefinder
                    contains similarity to Pfam domain: PF00069 (Eukaryotic
                    protein kinase domain), Score=322.1, E-value=2.1e-93, N=1
                    cDNA EST yk233b4.3 comes from this gene
                    cDNA EST yk233b4.5 comes from this gene
                    cDNA EST yk307d8.5 comes from this gene
                    cDNA EST yk451g9.5 comes from this gene
                    cDNA EST yk387f2.5 comes from this gene
                    cDNA EST yk337g9.5 comes from this gene
                    cDNA EST yk463c11.5 comes from this gene
                    cDNA EST yk507g2.5 comes from this gene
                    cDNA EST yk508c2.5 comes from this gene
                    cDNA EST yk520a5.5 comes from this gene
                    cDNA EST yk541h5.5 comes from this gene
                    cDNA EST yk590d9.5 comes from this gene
                    cDNA EST yk640e7.5 comes from this gene"
                    /codon_start=1
                    /protein_id="CAA22311.1"
                    /db_xref="GI:3979932"
                    /db_xref="SPTREMBL:Q9U2Q9"
                    /translation="MNKQLLSCSLKSGKQVTMVVASVATDGVDQQVEISYYDQKVIGN
                    GSFGVVFLAKLSTTNEMVAIKKVLQDKRFKNRELQIMRKLNHPNIVKLKYFFYSSGEK
                    KDELYLNLILEYVPETVYRVARHYSKQRQQIPMIYVKLYMYQLLRSLAYIHSIGICHR
                    DIKPQNLLIDPESGVLKLCDFGSAKYLVRNEPNVSYICSRYYRAPELIFGATNYTNSI
                    DVWSAGTVMAELLLGQPIFPGDSGVDQLVEIIKVLGTPTREQIQSMNPNYKEFKFPQI
                    KAHPWNKVFRVHTPAEAIDLISKIIEYTPTSRPTPQAACQHAFFDELRNPDARLPSGR
                    PLPTLEMDGPMGTGEVSTTSGDVAGPSA"
     gene           join(47525. .47593,48112. .48208,48610. .48751,49616. .50045,
                    50501. .50818,51553. .51971,52106. .52364,52847. .53224)
                    /gene="Y18D10A.6"
     CDS            join(47525. .47593,48112. .48208,48610. .48751,49616. .50045,
                    50501. .50818,51553. .51971,52106. .52364,52847. .53224)
                    /gene="Y18D10A.6"
```

```
                     /note="contains similarity to Pfam domain: PF00999
                     (Sodium/hydrogen exchanger family), Score=354.0,
                     E-value=5.3e-103, N=1"
                     /codon_start=1
                     /protein_id="CAA22320.1"
                     /db_xref="GI:3979941"
                     /db_xref="SPTREMBL:Q9XW14"
                     /translation="MMLSVEEQVLRNNIELMFTFNSLKILANGVTWQLWESTLNQGTA
                     TSGIMRFALKTALSICIFLLIFQTVDSDSSDSSASASVVSGAVKSEDTVVAVNKTDVL
                     GEAIDANATSLEQHGAAIVGNVSEEKKRSLAIFFILFVIMLATLVVHMLIVSKIHWMP
                     ESLAIVALGALIGSILSYSRRDWSEIEALSPDVFFLVLLPPIIFENAYNLNKGYFFSN
                     FVPILTFAIFGTTISAMVIGAGLYILGAIGLIFEEFTFFECFAFAAMISAVDPVGTLAI
                     FQAVKVESLLYMLVFGESMLNDAVSIVLAATALRHAKPSFNSLPASEIITSAFVTFTE
                     MFFFSACLGVGIGLLSALLFKHVDLRKTPSLEFALLLIFSYIPYGFAEALDLSGIMAI
                     LFCGISMSQFTRHNVSPIAQITFRHTFRTISFVAETSTFAYIGMAFFTIKLNFAPWLI
                     FWSVVLCLLGRACNVFPLAYLVNQCRKDVQISMKNQIIMWFSGMRGAVCFALVLYMDL
                     DKEKKSILLTTVLFLILFTTIFLGGSALPFISFINRCYPNERQRKRRRTPRNKESTGN
                     SSALMMSKTQEMSFFGSDDWGPKKSALDATSSAGRIMRQLFVRKFTAIERLENRDKLA
                     ALTKRALASDQMTDSDDVEFGGGGGVGGGGRMKDDVTPTRGRSGSRNSSDVIISAGGG
                     GVSGEHHLLISSGSDSSTNEF"

  Query Match              5.8%;  Score 56.4;  DB 6;  Length 152878;
  Best Local Similarity   44.9%;  Pred. No. 0.45;
  Matches  258;  Conservative    0;  Mismatches  311;  Indels    5;  Gaps    1;

Qy     195 taaatattttagaaatatgtacaaagggttttattacagccgcccagataaactagacaa 254
           ||||||  |   || ||  || | |      |||||         | | ||| ||   ||
Db    2285 TAAATAAATATTAATTAAATAAATAAATAAATATTAATTAAATAAATAAAAATTAATTAA 2344

Qy     255 atgtttttacaaatactatggacttaataagttaattaatttattgataaaagaaaagcc 314
           ||   |  | |||||  ||   |  |||  |  ||| || |   |  |||||  || |
Db    2345 ATAAATGAATAAATATTAATTAAATAAACAAATAAATATTAATTAAATAAATAAATAAAT 2404

Qy     315 agatttaatattattaacgtttcctacaccagttatgtcggtactaactgagcaatttaa 374
           |    | | || | |||   ||  |  |  ||  |   || ||| | |  || | ||
Db    2405 ATTAATTAAATAAATAAATATTAATTAAATAAATAAATAAATATTAATTAAATAAATAAA 2464

Qy     375 cattaatattccagttgctacagtgatgacagactatcgcttacataaaaactggattac 434
           ||||||     |  |  ||     || |   | |||   ||| ||||| |    | ||
Db    2465 AATTAATTAAATAAATATTAATTAAATAAATAAATATTAATTAGATAAATAAATAAATAT 2524

Qy     435 gccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagacttcatagacgt 494
           ||   ||  | |||||| |    | | ||  ||  |   |  |  | |  ||
Db    2525 TAATTAAATAAATAAATATTAATTAAATAAATAAACAAATATTAATTAAATAAATGAATA 2584

Qy     495 aggtattgatccttcaacagttaaagtgacaggtattcctattgataacaaatttgaaac 554
           |    |   |   | ||||  ||||        || |   ||  ||||  || |   ||
Db    2585 AATATTAATTAAATAAACAAATAAAT-----ATTAATTAAATAAATAAATAAATATTAAT 2639

Qy     555 gcctattaatcaaaagcagtggttaatagacaacaacttagatccagataagcaaactat 614
           |   | |   | |   |||   |  |  | ||| ||  | | |    || ||
Db    2640 TAAATAAATAAATATTAATTAAATAAATAAATATTAATTAAATAAATAAATATTAATTAG 2699

Qy     615 tttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatgattactgatat 674
           | |||          |  |||   |  ||  | |    || |   |  |    ||
Db    2700 ATAAATAAATAAATATTAATTAAATAAATAAATATTAATTAAATAAATAAACAAATATTA 2759

Qy     675 attagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagcaaagagctaaa 734
           ||||   ||| |   ||||    |  | | |  |  |        |  ||| |  ||||
Db    2760 ATTAAATAAATGAATAAATATTAATTAAATAAACAAATAAATATTAATTAAATAAATAAA 2819

Qy     735 gcgttctttaacagctaagtttaaattaacgaga 768
             | || |  |  ||| |  | |||||  | |
Db    2820 TAAATATTAATTAAATAAATAAATATTAATTAAA 2853


RESULT  15
CEY18D10
LOCUS       CEY18D10   247332 bp    DNA             HTG       07-SEP-1999
DEFINITION  Caenorhabditis elegans chromosome I clone Y18D10, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL008871
VERSION     AL008871.1  GI:3900862
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 247332)
  AUTHORS   Harris,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-1999) Nematode Sequencing Project, Sanger Centre,
            Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
            Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Nov 21, 1998 this sequence version replaced gi:3873430.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .247332
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="I"
                     /clone="Y18D10"
BASE COUNT    79342 a  43026 c  42968 g  79596 t   2400 others
ORIGIN


  Query Match              5.8%;  Score 56.4;  DB 83;  Length 247332;
  Best Local Similarity   44.9%;  Pred. No. 0.43;
  Matches  258;  Conservative    0;  Mismatches  311;  Indels    5;  Gaps    1;

Qy     195 taaatattttagaaatatgtacaaagggttttattacagccgcccagataaactagacaa 254
           ||||||  |   || ||  || | |      |||||         | | ||| ||   ||
Db   77282 TAAATAAATATTAATTAAATAAATAAATAAATATTAATTAAATAAATAAAAATTAATTAA 77341

Qy     255 atgtttttacaaatactatggacttaataagttaattaatttattgataaaagaaaagcc 314
           ||   |  | |||||  ||   |  |||  |  ||| || |   |  |||||  || |
Db   77342 ATAAATGAATAAATATTAATTAAATAAACAAATAAATATTAATTAAATAAATAAATAAAT 77401

Qy     315 agatttaatattattaacgtttcctacaccagttatgtcggtactaactgagcaatttaa 374
           |    | | || | |||   ||  |  |  |  ||  |   || ||| | |  || | ||
Db   77402 ATTAATTAAATAAATAAATATTAATTAAATAAATAAATAAATATTAATTAAATAAATAAA 77461

Qy     375 cattaatattccagttgctacagtgatgacagactatcgcttacataaaaactggattac 434
           ||||||     |  |  ||     || |   | |||   ||| ||||| |    | ||
Db   77462 AATTAATTAAATAAATATTAATTAAATAAATAAATATTAATTAGATAAATAAATAAATAT 77521

Qy     435 gccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagacttcatagacgt 494
           ||   ||  | |||||| |    | | ||  ||  |   |  |  | |  ||
Db   77522 TAATTAAATAAATAAATATTAATTAAATAAATAAACAAATATTAATTAAATAAATGAATA 77581

Qy     495 aggtattgatccttcaacagttaaagtgacaggtattcctattgataacaaatttgaaac 554
           |    |   |   | ||||  ||||        || |   ||  ||||  || |   ||
Db   77582 AATATTAATTAAATAAACAAATAAAT-----ATTAATTAAATAAATAAATAAATATTAAT 77636

Qy     555 gcctattaatcaaaagcagtggttaatagacaacaacttagatccagataagcaaactat 614
           |   | |   | |   |||   |  |  | ||| ||  | | |    || ||
Db   77637 TAAATAAATAAATATTAATTAAATAAATAAATATTAATTAAATAAATAAATATTAATTAG 77696

Qy     615 tttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatgattactgatat 674
           | |||          |  |||   |  ||  | |    || |   |  |    ||
Db   77697 ATAAATAAATAAATATTAATTAAATAAATAAATATTAATTAAATAAATAAACAAATATTA 77756

Qy     675 attagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagcaaagagctaaa 734
           ||||   ||| |   ||||    |  | | |  |  |        |  ||| |  ||||
Db   77757 ATTAAATAAATGAATAAATATTAATTAAATAAACAAATAAATATTAATTAAATAAATAAA 77816
```

```
Qy    735 gcgttctttaacagctaagtttaaattaacgaga 768
              | || |  |  ||| |  | |||||  | |
Db  77817 TAAATATTAATTAAATAAATAAATATTAATTAAA 77850
```

Search completed: June 29, 2001, 12:01:42
Job time: 8915 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 11:49:30 ; Search time 162.85 Seconds
(without alignments)
3759.313 Million cell updates/sec

Title: US-09-668-788-3
Perfect score: 975
Sequence: 1 atggttactcaaaataaaaa.....gtaaaacgctgatactccag 975

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	883.2	90.6	12658	18	AAV74438	Staphylococcus aur
c	2	437	44.8	2600	20	AAX08512	MurE gene. Staphy
c	3	182.2	18.7	936	22	AAF58252	Oligonucleotide D1
c	4	182.2	18.7	936	22	AAF58254	Oligonucleotide D1
c	5	182.2	18.7	936	22	AAF58257	Oligonucleotide D1
c	6	182.2	18.7	936	22	AAF58259	Oligonucleotide D2
c	7	182.2	18.7	936	22	AAF58262	Oligonucleotide D2
c	8	182.2	18.7	938	22	AAF58255	Oligonucleotide D1
	9	175.2	18.0	936	22	AAF58252	Oligonucleotide D1
	10	175.2	18.0	936	22	AAF58254	Oligonucleotide D1
	11	175.2	18.0	936	22	AAF58257	Oligonucleotide D1
	12	175.2	18.0	936	22	AAF58259	Oligonucleotide D2
	13	175.2	18.0	936	22	AAF58262	Oligonucleotide D2
	14	175.2	18.0	938	22	AAF58255	Oligonucleotide D1
	15	118	12.1	1149	20	AAZ23386	B. subtilis glycos
c	16	66	6.8	244	22	AAF58238	Oligonucleotide D1
	17	65.8	6.7	244	22	AAF58238	Oligonucleotide D1
	18	51.2	5.3	1764	21	AAC47190	Arabidopsis thalia
c	19	47.2	4.8	580073	18	AAT58840	Mycoplasma genital
	20	47	4.8	1266	19	AAV12734	Monogalactosyldiac
c	21	47	4.8	11011	21	AAC68252	B. burgdorferi tox
	22	46.4	4.8	8920	15	AAQ62924	Carbamoyl-phosphat
	23	45.8	4.7	11922	21	AAA70187	Plasmodium falcipa
	24	44.6	4.6	163319	21	AAF22306	Arabidopsis thalia
	25	43.8	4.5	19124	18	AAT72882	Plasmodium var-7 g
	26	43.8	4.5	19124	21	AAZ98287	Plasmodium var-7 p
	27	43.4	4.5	9789	17	AAT41852	cDNA encoding Plas
	28	43.2	4.4	32768	20	AAX13037	Enterococcus faeca
c	29	42.8	4.4	549	21	AAC95242	Cat flea head and
	30	42.8	4.4	1431	21	AAZ37082	DNA sequence encod
	31	42.8	4.4	1671	13	AAQ24134	50 kD subunit of S
c	32	42.6	4.4	116277	20	AAX20249	Borrelia burgdorfe
c	33	42.4	4.3	1416	21	AAA70135	Plasmodium falcipa
	34	42	4.3	3366	21	AAA70175	Plasmodium falcipa
c	35	42	4.3	4590	7	AAN60472	Sequence encoding
c	36	42	4.3	19124	18	AAT72882	Plasmodium var-7 g
c	37	42	4.3	19124	21	AAZ98287	Plasmodium var-7 p
	38	41.8	4.3	2486	21	AAA97037	Nucleotide sequenc
	39	41.4	4.2	6265	20	AAX08523	NBP46 (root lectin
	40	41	4.2	2058	19	AAV07560	Neocallimastix pat
	41	41	4.2	2058	20	AAZ11460	N. patriciarum xyl
	42	41	4.2	2058	22	AAC66514	N. patriciarum end
c	43	40.8	4.2	3369	21	AAA70230	Plasmodium falcipa
	44	40.6	4.2	4317	21	AAA70204	Plasmodium falcipa
	45	40.6	4.2	7143	21	AAA70250	Plasmodium falcipa

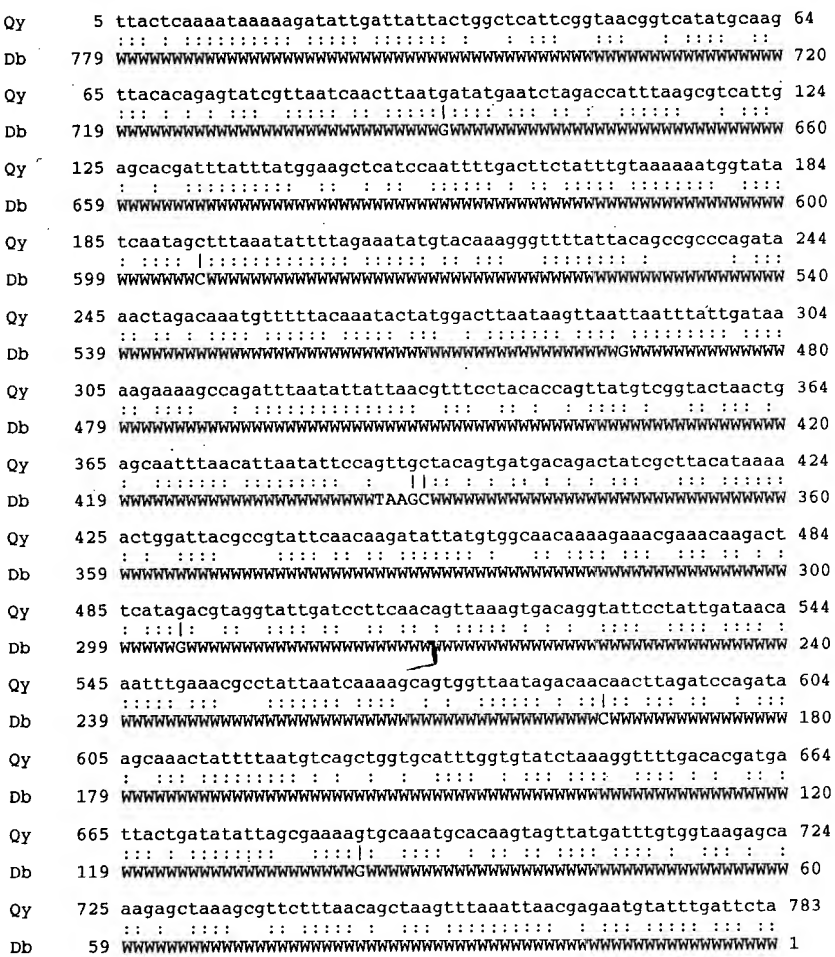
ALIGNMENTS

```
RESULT 1
AAV74438/c
ID   AAV74438 standard; DNA; 12658 BP.
XX
AC   AAV74438;
XX
DT   16-MAR-1999  (first entry)
XX
DE   Staphylococcus aureus contig SEQ ID #127.
XX
KW   Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW   cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW   skin infection; surgical wound infection; scalded skin syndrome;
KW   toxic shock syndrome; ds.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    181..240
FT                   /*tag= a
FT                   /note= "these bases represent a line of missing text in
FT                         the sequence listing in the specification. They
FT                         are included to maintain the nucleotide numbering
FT                         given in the specification for this DNA sequence"
FT   misc_feature    1981..2040
FT                   /*tag= b
FT                   /note= "these bases represent a line of missing text in
FT                         the sequence listing in the specification. They
FT                         are included to maintain the nucleotide numbering
FT                         given in the specification for this DNA sequence"
FT   misc_feature    3781..3840
FT                   /*tag= c
FT                   /note= "these bases represent a line of missing text in
FT                         the sequence listing in the specification. They
```

```
FT                                are included to maintain the nucleotide numbering
FT                                given in the specification for this DNA sequence"
FT   misc_feature        5581..5640
FT                       /*tag= d
FT                       /note= "these bases represent a line of missing text in
FT                              the sequence listing in the specification. They
FT                              are included to maintain the nucleotide numbering
FT                              given in the specification for this DNA sequence"
FT   misc_feature        7381..7440
FT                       /*tag= e
FT                       /note= "these bases represent a line of missing text in
FT                              the sequence listing in the specification. They
FT                              are included to maintain the nucleotide numbering
FT                              given in the specification for this DNA sequence"
FT   misc_feature        9181..9240
FT                       /*tag= f
FT                       /note= "these bases represent a line of missing text in
FT                              the sequence listing in the specification. They
FT                              are included to maintain the nucleotide numbering
FT                              given in the specification for this DNA sequence"
FT   misc_feature        10981..11040
FT                       /*tag= g
FT                       /note= "these bases represent a line of missing text in
FT                              the sequence listing in the specification. They
FT                              are included to maintain the nucleotide numbering
FT                              given in the specification for this DNA sequence"
XX
PN   EP786519-A2.
XX
PD   30-JUL-1997.
XX
PF   07-JAN-1997;   97EP-0100117.
XX
PR   05-JAN-1996;   96US-0009861.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Fannon MR,  Kunsch CA;
PI   Rosen CA;
XX
DR   WPI; 1997-374922/35.
XX
PT   Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT   stored on computer readable medium and used in the production of
PT   anti-S.aureus vaccines
XX
PS   Claim 1; Page 692-699; 3271pp; English.
XX
CC   This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC   of the invention. The DNA sequences are recorded on a computer readable
CC   medium, preferably selected from a floppy or hard disk, random access
CC   memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC   the S.aureus DNA sequences allows putative functions to be assigned so
CC   that protein-encoding or regulatory regions of commercial, therapeutic or
CC   industrial importance can be obtained. Specifically, sequences which are
CC   likely to encode antigens have been identified and these polypeptides can
CC   be used in a vaccine composition against S.aureus infection. The
CC   polypeptides can also be used in a kit for the immunodetection of
CC   S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC   including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC   skin and surgical wound infections, scalded skin syndrome, toxic shock
CC   syndrome, etc. Organisms transformed with the DNA sequences can be used
CC   for recombinant production of the polypeptides. The new DNA sequences
CC   (and their fragments) are useful as primers or probes for isolating
CC   homologues of any of the S.aureus DNA sequences contained on the
CC   computer readable medium.
XX
SQ   Sequence 12658 BP; 4296 A; 1711 C; 2283 G; 3942 T; 426 other;

  Query Match             90.6%;  Score 883.2;  DB 18;  Length 12658;
  Best Local Similarity   92.4%;  Pred. No. 1.5e-183;
  Matches  902;  Conservative    0;  Mismatches   73;  Indels    1;  Gaps    1;

Qy          1 atggttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatg 60
              ||||||||||||
Db       5652 ATGGTTACTCAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5593

Qy         61 caagttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtc 120
                          ||||||||||||||||||||||||||||||||||||||||||||||||
Db       5592 NNNNNNNNNNNNAGTATCGTTAATCAACTTAATGATATGAATCTAGACCATTTAAGCGTC 5533

Qy        121 attgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatgg 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5532 ATTGAGCACGATTTATTTATGGAAGCTCATCCAATTTTGACTTCTATTTGTAAAAAATGG 5473

Qy        181 tatatcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgccca 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5472 TATATCAATAGCTTTAAATATTTTAGAAATATGTACAAAGGGTTTTATTACAGCCGCCCA 5413

Qy        241 gataaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattg 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5412 GATAAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTTATTG 5353

Qy        301 ataaaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtacta 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5352 ATAAAAGAAAAGCCAGATTTAATATTATTAACGTTTCCTACACCAGTTATGTCGGTACTA 5293

Qy        361 actgagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacat 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5292 ACTGAGCAATTTAACATTAATATTCCAGTTGCTACAGTGATGACAGACTATCGCTTACAT 5233

Qy        421 aaaaactggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaa 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5232 AAAAACTGGATTACGCCGTATTCAACAAGATATTATGTGGCAACAAAAGAAACGAAACAA 5173

Qy        481 gacttcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgat 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5172 GACTTCATAGACGTAGGTATTGATCCTTCAACAGTTAAAGTGACAGGTATTCCTATTGAT 5113

Qy        541 aacaaatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatcca 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5112 AACAAATTTGAAACGCCTATTAATCAAAAGCAGTGGTTAATAGACAACAACTTAGATCCA 5053

Qy        601 gataagcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacg 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5052 GATAAGCAAACTATTTTAATGTCAGCTGGTGCATTTGGTGTATCTAAAGGTTTTGACACG 4993

Qy        661 atgattactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaag 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4992 ATGATTACTGATATATTAGCGAAAAGTGCAAATGCACAAGTAGTTATGATTTGTGGTAAG 4933

Qy        721 agcaaagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgatt 780
              ||||||||||||||||||||||||||||||||||||||     |       |      ||||||
Db       4932 AGCAAAGAGCTAAAGCGTTCTTTAACAGCTAAGTTTAAATCTAACGAGAATGTATTGATT 4873

Qy        781 ctaggttataccaaacacatgaatgaatggatggcatcaagtcaacttatgattacgaaa 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4872 CTAGGTTATACCAAACACATGAATGAATGGATGGCATCAAGTCAACTTATGATTACGAAA 4813

Qy        841 cctggtggtatcacaataactgaaggtttcgcccgttgtattccaatgattttcctaaat 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4812 CCTGGTGGTATCACAATAACTGAAGGTTTCGCCCGTTGTATTCCAATGATTTTCCTAAAT 4753

Qy        901 cctgcacctggtcaagagcttgaaaatgccttttactttgaagaaaaaggttttggtaaa 960
              |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db       4752 CCTGCACCTGGTCAAGAGCTTGAAAATGCCCTTTACTTTGAAGAAAAAGGTTTTGGTAAA 4693

Qy        961 a-cgctgatactccag 975
              | ||||||||||||||
Db       4692 ATCGCTGATACTCCAG 4677

RESULT  2
AAX08512/c
ID   AAX08512 standard; DNA; 2600 BP.
```

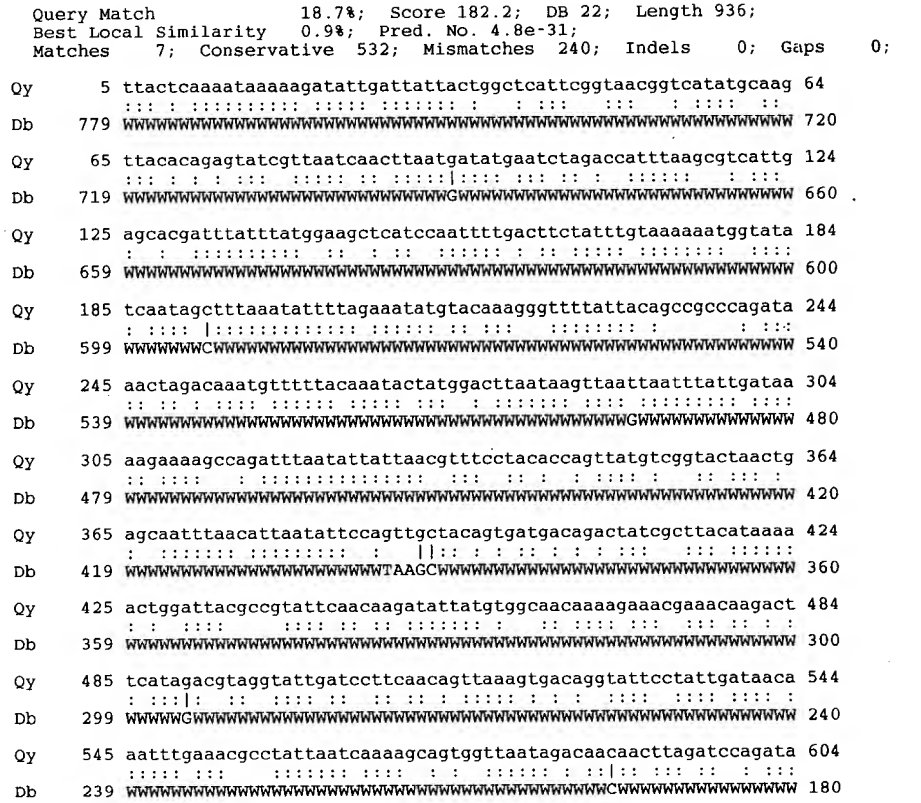
```
XX
AC   AAX08512;
XX
DT   19-JUL-1999  (first entry)
XX
DE   MurE gene.
XX
KW   MurE; Staphylococcus aureus; infection; treatment; diagnosis;
KW   primer; osteomyelitis; septic arthritis; septic thrombophlebitis;
KW   PCR; scalded skin syndrome; antibody; antibiotic; ds.
XX
OS   Staphylococcus aureus.
XX
FH   Key             Location/Qualifiers
FT   CDS             993..2473
FT                   /*tag= a
FT                   /product= "UDP-N-acetylmuramoyl-L-alanyl-D-glutamate:
FT                            lysine ligase"
XX
PN   EP905248-A1.
XX
PD   31-MAR-1999.
XX
PF   21-SEP-1998;   98EP-0307624.
XX
PR   10-DEC-1997;   97US-0988251.
PR   25-SEP-1997;   97US-0059956.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Shilling LK,  Wallis NG,  Zabierowski S;
XX
DR   WPI; 1999-206775/18.
DR   P-PSDB; AAW85670.
XX
PT   New MurE gene from Staphylococcus aureus useful for diagnosing and
PT   treating bacterial diseases such as osteomyelitis and toxic shock
PT   syndrome
XX
PS   Claim 6; Page 5-8; 32pp; English.
XX
CC   Staphylococcus aureus can cause both invasive and toxigenic
CC   infections which lead to diseases such as osteomyelitis, septic
CC   arthritis, septic thrombophlebitis, scalded skin syndrome and toxic
CC   shock syndrome. Antagonists of MurE such as antibodies may be used
CC   to treat individuals with these diseases and as a prophylactic
CC   against them by bathing wounds and coating implanted devices.
CC   Antagonists of MurE may also be used to treat diseases caused by
CC   Helicobacter pylori infection such as stomach ulcers and
CC   gastrointestinal carcinoma. This method of treatment of bacterial
CC   infection is advantageous since many strains of Staphylococcus
CC   aureus are now resistant to some or all of the standard antibiotics.
CC   MurE is an important enzyme in bacterial metabolism and so the
CC   identification of such an enzyme in Staphylococcus aureus can be
CC   used to effectively diagnose and treat diseases caused by this
CC   pathogen.  Detection of murE, and hence a bacterial infection,
CC   can be done by reverse transcriptase polymerase chain reaction
CC   (RT-PCR), preferably using the primers described in AAX08514, AAX08515.
XX
SQ   Sequence 2600 BP; 893 A; 370 C; 493 G; 844 T; 0 other;


  Query Match             44.8%;  Score 437;  DB 20;  Length 2600;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-86;
  Matches  437;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatg 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     557 ATGGTTACTCAAAATAAAAAGATATTGATTATTACTGGCTCATTCGGTAACGGTCATATG 498

Qy      61 caagttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     497 CAAGTTACACAGAGTATCGTTAATCAACTTAATGATATGAATCTAGACCATTTAAGCGTC 438

Qy     121 attgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatgg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     437 ATTGAGCACGATTTATTTATGGAAGCTCATCCAATTTTGACTTCTATTTGTAAAAAATGG 378

Qy     181 tatatcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgccca 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     377 TATATCAATAGCTTTAAATATTTTAGAAATATGTACAAAGGGTTTTATTACAGCCGCCCA 318

Qy     241 gataaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattg 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     317 GATAAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTTATTG 258

Qy     301 ataaaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtacta 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     257 ATAAAAGAAAAGCCAGATTTAATATTATTAACGTTTCCTACACCAGTTATGTCGGTACTA 198

Qy     361 actgagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacat 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 ACTGAGCAATTTAACATTAATATTCCAGTTGCTACAGTGATGACAGACTATCGCTTACAT 138

Qy     421 aaaaactggattacgcc 437
           |||||||||||||||||
Db     137 AAAAACTGGATTACGCC 121


RESULT   3
AAF58252/c
ID   AAF58252 standard; DNA; 936 BP.
XX
AC   AAF58252;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1835.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000;  2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;
```

Query Match 18.7%; Score 182.2; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 4.8e-31;
Matches 7; Conservative 532; Mismatches 240; Indels 0; Gaps 0;



RESULT 4
AAF58254/c
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 18.7%; Score 182.2; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 4.8e-31;
Matches 7; Conservative 532; Mismatches 240; Indels 0; Gaps 0;



```
Qy    605 agcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatga 664
          :  ::: :::::::::: : :  :  :  :::: : ::: ::::  :::: : :  :: :
Db    179 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 120

Qy    665 ttactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagca 724
          ::: : ::::::::   ::::|:  ::::  : :: :: :::: :::: :  ::: :  :
Db    119 WWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 60

Qy    725 aagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgattcta 783
          :: :  ::::   :: ::::: :  ::: :::::::::  : ::: ::::: ::: ::
Db     59 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 1
```

RESULT 5
AAF58257/c
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1954.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

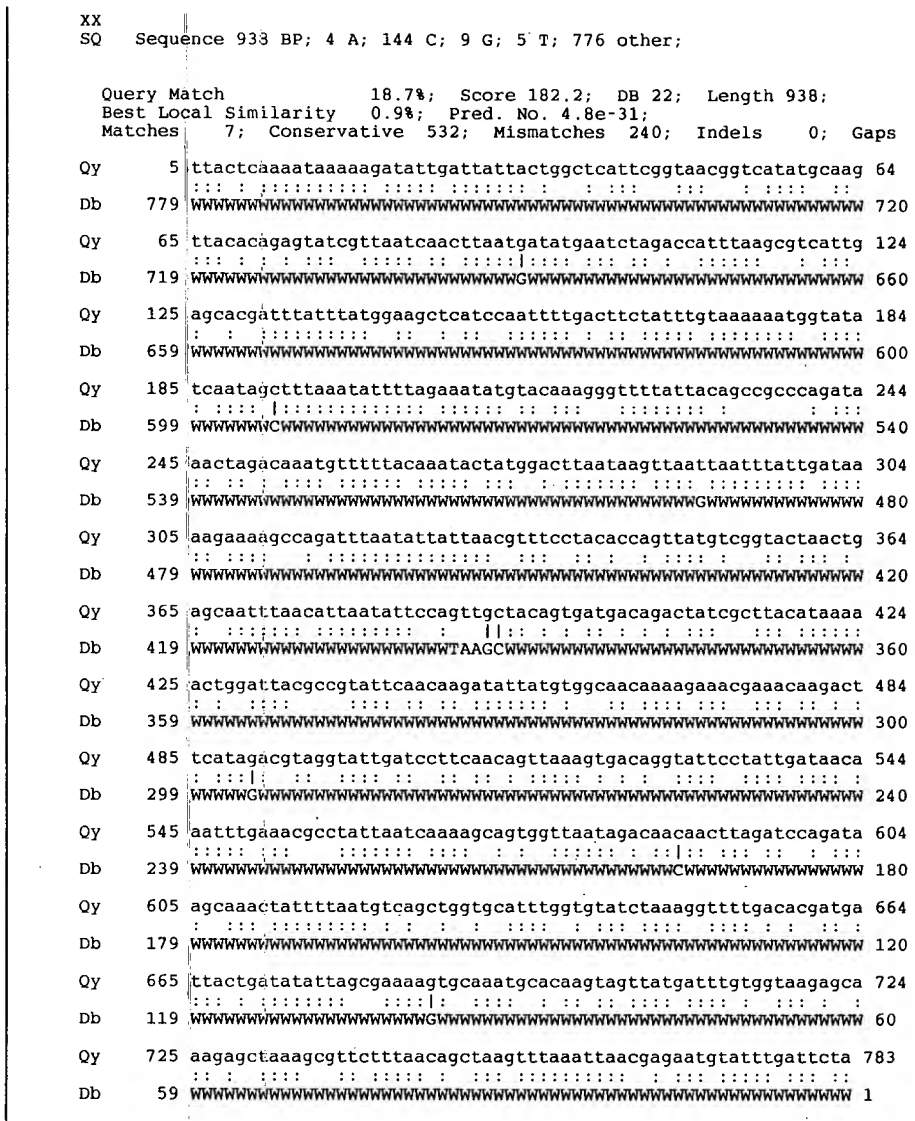
Query Match 18.7%; Score 182.2; DB 22; Length 936;
Best Local Similarity 0.9%; Pred. No. 4.8e-31;
Matches 7; Conservative 532; Mismatches 240; Indels 0; Gaps 0;

```
Qy      5 ttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatgcaag 64
          ::: : :::::::::: ::::: ::::::: :   : :::  :::   : :::: ::
Db    779 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 720

Qy     65 ttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattg 124
          ::: : : : :::  ::::: :: :::::|:::: ::: :: :  ::::::   : :::
Db    719 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWW 660

Qy    125 agcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtata 184
          :  :  ::::::::::  ::  : ::  :::::: : :: ::::: :::::::  ::::
Db    659 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 600

Qy    185 tcaatagctttaaatattttagaaatatgtacaaaggttttattacagccgcccagata 244
          : :::: |:::::::::::::: :::::: :: :::   :::::::: :        : :::
Db    599 WWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 540

Qy    245 aactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgataa 304
          :: :: : :::: :::::: ::::: :::  : ::::::: :::: :::::::::: ::::
Db    539 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWW 480

Qy    305 aagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactg 364
          :: ::::   : ::::::::::::::::  :::  :: :  : :::: :   :: ::: :
Db    479 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 420

Qy    365 agcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataaaa 424
          :  ::::::: :::::::::  :   ||:: : : :: : : : :::   ::: ::::::
Db    419 WWWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 360

Qy    425 actggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagact 484
          : :  ::::     :::: :: :: :::::::: :   :: :::: ::: ::: :: : :
Db    359 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 300

Qy    485 tcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataaca 544
          : :::|:  ::  :::: ::  :: :: : ::::: : : :  :::: :::: :::: :
Db    299 WWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 240

Qy    545 aatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccagata 604
          ::::: :::    ::::::: :::  : :  :::::: : ::|:: ::: ::  : :::
Db    239 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWWWWWWW 180

Qy    605 agcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatga 664
          :  ::: :::::::::: : :  :  :  :::: : ::: ::::  :::: : :  :: :
Db    179 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 120

Qy    665 ttactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagca 724
          ::: : ::::::::   ::::|:  ::::  : :: :: :::: :::: :  ::: :  :
Db    119 WWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 60

Qy    725 aagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgattcta 783
          :: :  ::::   :: ::::: :  ::: :::::::::  : ::: ::::: ::: ::
Db     59 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 1
```

RESULT 6
AAF58259/c
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.

```
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface -
XX
PS   Example 6; Page 128; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

  Query Match             18.7%;  Score 182.2;  DB 22;  Length 936;
  Best Local Similarity   0.9%;  Pred. No. 4.8e-31;
  Matches    7;  Conservative  532;  Mismatches  240;  Indels    0;  Gaps    0;

Qy      5 ttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatgcaag 64
          ::: : :::::::::: ::::: ::::::: :   : :::   :::   : ::::  ::
Db    779 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 720

Qy     65 ttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattg 124
          ::: : : : :::  ::::: :: :::::|:::: ::: :: :  ::::::   : :::
Db    719 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWW 660

Qy    125 agcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtata 184
          :  :  :::::::::  ::  : ::  :::::: : :: ::::: :::::::  ::::
Db    659 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 600

Qy    185 tcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagata 244
          : :::: |:::::::::::: :::::: :: :::   :::::::: :       : :::
Db    599 WWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 540

Qy    245 aactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgataa 304
          :: :: : :::: :::::: ::::: ::::  : :::::::: :::: :::::::::: ::::
Db    539 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWW 480

Qy    305 aagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactg 364
          :: ::::   : ::::::::::::::  :::  :: :  : :::: :   :: ::: :
Db    479 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 420

Qy    365 agcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataaaa 424
          :  ::::::: :::::::::  :   ||:: : : :: : : : :  :::  :::::::
Db    419 WWWWWWWWWWWWWWWWWWWWWWWTAAGCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 360

Qy    425 actggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagact 484
          : :  ::::     :::: :: :: :::::::: :   :: :::: :::  ::: :: : :
Db    359 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 300

Qy    485 tcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataaca 544
          : :::|:  ::  :::: ::  :: :: : ::::: : : :  ::::  :::: :::: :
Db    299 WWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 240

Qy    545 aatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccagata 604
          ::::: :::    ::::::: :::: : :  :::::: : ::|:: ::: ::  : :::
Db    239 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWCWWWWWWWWWWWW 180

Qy    605 agcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatga 664
          :  ::: :::::::::: : :  :  :  :::: : ::: :::: :::: : :  :: :
Db    179 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 120

Qy    665 ttactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagca 724
          ::: : ::::::::   ::::|:  ::::  : :: :: :::: :::: :  ::: :  :
Db    119 WWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 60

Qy    725 aagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgattcta 783
          :: :  ::::   :: ::::: :  ::: ::::::::::  : ::: ::::: ::: ::
Db     59 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 1

RESULT   7
AAF58262/c
ID   AAF58262 standard; DNA; 936 BP.
XX
AC   AAF58262;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D2007.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface -
XX
PS   Example 6; Page 128; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

  Query Match             18.7%;  Score 182.2;  DB 22;  Length 936;
  Best Local Similarity   0.9%;  Pred. No. 4.8e-31;
  Matches    7;  Conservative  532;  Mismatches  240;  Indels    0;  Gaps    0;

Qy      5 ttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatgcaag 64
          ::: : :::::::::: ::::: ::::::: :   : :::   :::   : ::::  ::
Db    779 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 720

Qy     65 ttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattg 124
          ::: : : : :::  ::::: :: :::::|:::: ::: :: :  ::::::   : :::
Db    719 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWWWWWWWWWWWWWWWWW 660

Qy    125 agcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtata 184
          :  :  :::::::::  ::  : ::  :::::: : :: ::::: :::::::  ::::
Db    659 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 600

Qy    185 tcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagata 244
          .: :::: |:::::::::::: :::::: :: :::   :::::::: :       : :::
Db    599 WWWWWWWCWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWW 540

Qy    245 aactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgataa 304
          :: :: : :::: :::::: ::::: ::::  : :::::::: :::: :::::::::: ::::
Db    539 WWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWGWWWWWWWWWWWWW 480

Qy    305 aagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactg 364
          :: ::::   : ::::::::::::::  :::  :: :  : :::: :   :: ::: :
```

RESULT 8
AAF58255/c
ID AAF58255 standard; DNA; 938 BP.
XX
AC AAF58255;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1876.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.



RESULT 9
AAF58252
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.

XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 18.0%; Score 175.2; DB 22; Length 936;
Best Local Similarity 0.4%; Pred. No. 1.6e-29;
Matches 3; Conservative 532; Mismatches 245; Indels 0; Gaps 0;

RESULT 10
AAF58254
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 18.0%; Score 175.2; DB 22; Length 936;
Best Local Similarity 0.4%; Pred. No. 1.6e-29;
Matches 3; Conservative 532; Mismatches 245; Indels 0; Gaps 0;

```
              ::: : : : ::: ::::: :: ::::: :::: ::: :: :  ::::::    : :::
Db        69 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 128

Qy       125 agcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtata 184
              :  :  ::::::::::  ::   : ::  :::::: : :: ::::: ::::::::  ::::
Db       129 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 188

Qy       185 tcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagata 244
              : ::::  :::::::::::: :::::: :: :::    ::::::::  :         : :::
Db       189 wwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 248

Qy       245 aactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgataa 304
              :: :: : :::: :::::: ::::: :::   : ::::::: :::: ::::::::: ::::
Db       249 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwww 308

Qy       305 aagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactg 364
             .:: ::::   : ::::::::::::::  :::   :: :   : :::: :    :: ::: :
Db       309 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 368

Qy       365 agcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataaaa 424
              :  ::::::: ::::::::  :| |  :: : : :: : : : :::    ::: ::::::
Db       369 wwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 428

Qy       425 actggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagact 484
              : :  ::::       :::: :: :: ::::::: :    :: :::: :::   ::: :: : :
Db       429 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 488

Qy       485 tcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataaca 544
              : ::: :   ::   :::: ::   :: :: : ::::: : : :    ::::   :::: :::: :
Db       489 wwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 548

Qy       545 aatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccagata 604
              ::::: :::     ::::::: ::::   :  :   :::::: : :: :: ::: ::   : :::
Db       549 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwwwwwwww 608

Qy       605 agcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatga 664
              :  ::: ::::::::: :  :   :   :   ::::   : ::: ::::   :::: :  :  :: :
Db       609 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 668

Qy       665 ttactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagca 724
              ::: : ::::::::    :::: :  ::::   : :: :: :::: :::: :  ::: :  :
Db       669 wwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 728

Qy       725 aagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgattctag 784
              :: :   ::::    :: ::::: :    ::: ::::::::::   : ::: ::::: ::: ::|
Db       729 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwg 788
```

RESULT 11
AAF58257
ID   AAF58257 standard; DNA; 936 BP.
XX
AC   AAF58257;
XX
DT   24-APR-2001 (first entry)
XX
DE   Oligonucleotide D1954.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999; 99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

```
  Query Match             18.0%;  Score 175.2;  DB 22;  Length 936;
  Best Local Similarity   0.4%;  Pred. No. 1.6e-29;
  Matches    3;  Conservative  532;  Mismatches  245;  Indels    0;  Gaps    0;

Qy         5 ttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatgcaag 64
              ::: : :::::::::: ::::: :::::::  :     : :::    :::    : ::::  ::
Db         9 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 68

Qy        65 ttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattg 124
              ::: : : : :::   ::::: :: ::::: :::: ::: :: :   ::::::    : :::
Db        69 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwww 128

Qy       125 agcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtata 184
              :  :  ::::::::::  ::   : ::  :::::: : :: ::::: ::::::::  ::::
Db       129 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 188

Qy       185 tcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagata 244
              : ::::  :::::::::::: :::::: :: :::    ::::::::  :         : :::
Db       189 wwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 248

Qy       245 aactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgataa 304
              :: :: : :::: :::::: ::::: :::   : ::::::: :::: ::::::::: ::::
Db       249 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwww 308

Qy       305 aagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactg 364
              :: ::::   : ::::::::::::::  :::   :: :   : :::: :    :: ::: :
Db       309 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 368

Qy       365 agcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataaaa 424
              :  ::::::: ::::::::  :| |  :: : : :: : : : :::    ::: ::::::
Db       369 wwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 428

Qy       425 actggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagact 484
              : :  ::::       :::: :: :: ::::::: :    :: :::: :::   ::: :: : :
Db       429 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 488

Qy       485 tcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataaca 544
              : ::: :   ::   :::: ::   :: :: : ::::: : : :    ::::   :::: :::: :
Db       489 wwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 548

Qy       545 aatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccagata 604
              ::::: :::     ::::::: ::::   :  :   :::::: : :: :: ::: ::   : :::
Db       549 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwwwwwwww 608

Qy       605 agcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatga 664
              :  ::: ::::::::: :  :   :   :   ::::   : ::: ::::   :::: :  :  :: :
Db       609 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 668

Qy       665 ttactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagca 724
              ::: : ::::::::    :::: :  ::::   : :: :: :::: :::: :  ::: :  :
Db       669 wwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 728
```

```
Qy    725 aagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgattctag 784
          :: :  ::::   :: ::::: :  ::: ::::::::::  : ::: ::::: ::: ::|
Db    729 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwg 788


RESULT 12
AAF58259
ID   AAF58259 standard; DNA; 936 BP.
XX
AC   AAF58259;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D2004.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface -
XX
PS   Example 6; Page 128; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

  Query Match          18.0%;  Score 175.2;  DB 22;  Length 936;
  Best Local Similarity  0.4%;  Pred. No. 1.6e-29;
  Matches    3;  Conservative  532;  Mismatches  245;  Indels    0;  Gaps    0;

Qy      5 ttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatgcaag 64
          ::: : :::::::::: ::::: :::::::: :   : :::   :::   : ::::  ::
Db      9 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 68

Qy     65 ttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattg 124
          ::: : : : :::  ::::: :: ::::: :::: ::: :: :  :::::   : :::
Db     69 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwCWwwwwwwwwwwwwwwwwwwwwwwwww 128

Qy    125 agcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtata 184
          :  :  ::::::::::  ::  : ::  :::::: : :: ::::: ::::::::  ::::
Db    129 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 188

Qy    185 tcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagata 244
          : ::::  :::::::::::::: :::::: :: :::   :::::::: :       : :::
Db    189 wwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 248

Qy    245 aactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgataa 304
          :: :: : :::: :::::: ::::: :::  : ::::::: :::: ::::::::: ::::

Db    249 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 308

Qy    305 aagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactg 364
          :: ::::   : ::::::::::::::::  :::  :: :  : :::: :   :: ::: :
Db    309 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 368

Qy    365 agcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataaaa 424
          :  ::::::: ::::::::: . :| |  :: : : :: : : : :::   ::: ::::::
Db    369 wwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwwww 428

Qy    425 actggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagact 434
          : :  ::::     :::: :: :: ::::::: :   :: :::: :::  ::: :: : :
Db    429 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 488

Qy    485 tcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataaca 544
          : ::: :  ::  :::: ::  :: :: : ::::: : : :  :::: :::: :::: :
Db    489 wwwwwCwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 548

Qy    545 aatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccagata 604
          ::::: :::    ::::::: :::: : :  :::::: : :: :: ::: ::  : :::
Db    549 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwww 608

Qy    605 agcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatga 664
          : ::: ::::::::: : :  :  :  :::: : ::: :::  :::: : :  :: :
Db    609 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 668

Qy    665 ttactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagca 724
          ::: : ::::::::   :::: :  ::::  : :: :: :::: :::: :  ::: :  :
Db    669 wwwwwwwwwwwwwwwwwwwwwwCwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 728

Qy    725 aagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgattctag 784
          :: :  ::::   :: ::::: :  ::: ::::::::::  : ::: ::::: ::: ::|
Db    729 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwg 788


RESULT 13
AAF58262
ID   AAF58262 standard; DNA; 936 BP.
XX
AC   AAF58262;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D2007.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface -
XX
PS   Example 6; Page 128; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
```

```
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 936 BP; 5 A; 139 C; 10 G; 6 T; 776 other;

  Query Match             18.0%;  Score 175.2;  DB 22;  Length 936;
  Best Local Similarity   0.4%;  Pred. No. 1.6e-29;
  Matches    3;  Conservative  532;  Mismatches  245;  Indels    0;  Gaps    0;

Qy        5 ttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatgcaag 64
             ::: : :::::::::: ::::: ::::::: :    : :::    :::   : ::::  ::
Db        9 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 68

Qy       65 ttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattg 124
             ::: : : : :::  ::::: :: ::::: :::: ::: :: :  ::::::    : :::
Db       69 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwww 128

Qy      125 agcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtata 184
             :  :  :::::::::  ::   : ::  :::::: : :: ::::: ::::::::   ::::
Db      129 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 188

Qy      185 tcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagata 244
             : ::::  ::::::::::::: :::::: :: :::    :::::::: :          : :::
Db      189 wwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 248

Qy      245 aactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgataa 304
             :: :: : :::: :::::: ::::: :::   : ::::::: :::: ::::::::: ::::
Db      249 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwww 308

Qy      305 aagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactg 364
             :: ::::    : ::::::::::::::   :::   :: :   : :::: :    :: ::: :
Db      309 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 368

Qy      365 agcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataaaa 424
             :  ::::::: ::::::::  :| |  :: : : : : : : : :::    ::: ::::::
Db      369 wwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwww 428

Qy      425 actggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagact 484
             : :  ::::        :::: :: :: ::::::: :    :: :::: :::   ::: :: : :
Db      429 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 488

Qy      485 tcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataaca 544
             : ::: :   ::   :::: ::   :: :: : ::::: : :  ::::   :::: :::: :
Db      489 wwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 548

Qy      545 aatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccagata 604
             ::::: :::      ::::::: ::::   : :   :::::: : :: :: :: :  : :::
Db      549 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwwwwww 608

Qy      605 agcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatga 664
             :   ::: ::::::::: : :   :   :   ::::   : ::: ::::   :::: : :  : :
Db      609 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 668

Qy      665 ttactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagca 724
             ::: : ::::::::    :::: :   ::::   : :: :: :::: :::: :    ::: :  :
Db      669 wwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 728

Qy      725 aagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgattctag 784
             :: :   ::::    :: :::::  :   ::: ::::::::::   : ::: ::::: ::: ::|
Db      729 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwg 788

RESULT 14
AAF58255
ID   AAF58255 standard; DNA; 938 BP.
XX
AC   AAF58255;
XX
DT   24-APR-2001  (first entry)
XX
DE   Oligonucleotide D1876.
XX
KW   Electron-transfer group; ETM; mismatch; genotyping;
KW   gene expression; ss.
XX
OS   Synthetic.
XX
PN   WO200107665-A2.
XX
PD   01-FEB-2001.
XX
PF   26-JUL-2000; 2000WO-US20476.
XX
PR   26-JUL-1999;   99US-0145695.
PR   17-MAR-2000; 2000US-0190259.
XX
PA   (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI   Umek RM;
XX
DR   WPI; 2001-159728/16.
XX
PT   Nucleic acids containing electron-transfer group, useful as labels in
PT   hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT   a single surface  -
XX
PS   Example 6; Page 127; 159pp; English.
XX
CC   The present invention relates to a composition comprising two nucleic
CC   acids each containing an electron-transfer group (ETM) having
CC   different redox potentials. The invention is used for electronic
CC   detection of nucleic acids, especially of substitutions (mismatches)
CC   and single-nucleotide polymorphisms, e.g. for genotyping,
CC   monitoring gene expression.
XX
SQ   Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

  Query Match             18.0%;  Score 175.2;  DB 22;  Length 938;
  Best Local Similarity   0.4%;  Pred. No. 1.6e-29;
  Matches    3;  Conservative  532;  Mismatches  245;  Indels    0;  Gaps    0;

Qy        5 ttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatgcaag 64
             ::: : :::::::::: ::::: ::::::: :    : :::    :::   : ::::  ::
Db        9 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 68

Qy       65 ttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattg 124
             ::: : : : :::  ::::: :: ::::: :::: ::: :: :  ::::::    : :::
Db       69 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwww 128

Qy      125 agcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtata 184
             :  :  :::::::::  ::   : ::  :::::: : :: ::::: ::::::::   ::::
Db      129 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 188

Qy      185 tcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagata 244
             : ::::  ::::::::::::: :::::: :: :::    :::::::: :          : :::
Db      189 wwwwwwwgwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 248

Qy      245 aactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgataa 304
             :: :: : :::: :::::: ::::: :::   : ::::::: :::: ::::::::: ::::
Db      249 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwww 308

Qy      305 aagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactg 364
             :: ::::    : ::::::::::::::   :::   :: :   : :::: :    :: ::: :
Db      309 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 368

Qy      365 agcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataaaa 424
             :  ::::::: ::::::::  :| |  :: : : : : : : : :::    ::: ::::::
Db      369 wwwwwwwwwwwwwwwwwwwwwwwwwgcttawwwwwwwwwwwwwwwwwwwwwwwwwwwwww 428

Qy      425 actggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagact 484
             : :  ::::        :::: :: :: ::::::: :    :: :::: :::   ::: :: : :
Db      429 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 488
```

```
Qy    485 tcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataaca 544
          : ::: :  ::  :::: ::  :: :: : ::::: : : :  :::: :::: :::: :
Db    489 wwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 548

Qy    545 aatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccagata 604
          ::::: :::    ::::::: :::: : :  :::::: : :: :: ::: ::  : :::
Db    549 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwgwwwwwwwwwwwwwww 608

Qy    605 agcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatga 664
          :  ::: ::::::::: : :  :  :  :::  : ::: ::::  :::: : :  :: :
Db    609 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 668

Qy    665 ttactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagca 724
          ::: : ::::::::   :::: :  ::::  : :: :: :::: :::: :  ::: :  :
Db    669 wwwwwwwwwwwwwwwwwwwwwwwcwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww 728

Qy    725 aagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttgattctag 784
          :: :  ::::   :: ::::: :  ::: :::::::::  : ::: ::::: ::: ::|
Db    729 wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwg 788
```

RESULT 15
AAZ23386
ID AAZ23386 standard; DNA; 1149 BP.
XX
AC AAZ23386;
XX
DT 13-DEC-1999 (first entry)
XX
DE B. subtilis glycosyl transferase catalytic domain DNA.
XX
KW Catalytic domain; glycosyl transferase; processive activity; detergent;
KW oligosaccharide glycolipid; 3-oligoglucosyl-1,2-diacylglycerol;
KW food emulsifier; polymer modifier; glycosyl transferase; ss.
XX
OS Bacillus subtilis.
XX
PN DE19819958-A1.
XX
PD 30-SEP-1999.
XX
PF 05-MAY-1998; 98DE-1019958.
XX
PR 25-MAR-1998; 98DE-1013017.
XX
PA (GVSE-) GVS GES ERWERB & VERW LANDWIRTSCHAFTLICH.
PA (BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.
XX
PI Wolter FP, Jorasch P, Heinz E, Zaehringer U;
XX
DR WPI; 1999-552364/47.
DR P-PSDB; AAY33444.
XX
PT New glycosyl transferase protein, useful for producing glycolipids -
XX
PS Example 1; Page 6; 8pp; German.
XX
CC This invention describes a novel protein (I) comprising identical or
CC different catalytic domains of glycosyl transferases and which has
CC processive activity (i.e. builds up oligosaccharide glycolipids by
CC successive addition of glucose units). (I) can be used to produce
CC 3-oligoglucosyl-1,2-diacylglycerols useful as food emulsifiers, polymer
CC modifiers or detergents. This sequence encodes a Bacillus subtilis
CC glycosyl transferase catalytic domain which is used in the method of
CC the invention.
XX
SQ Sequence 1149 BP; 380 A; 217 C; 244 G; 308 T; 0 other;

Query Match 12.1%; Score 118; DB 20; Length 1149;
Best Local Similarity 48.5%; Pred. No. 4.7e-17;
Matches 458; Conservative 0; Mismatches 465; Indels 21; Gaps 4;

```
Qy     13 aataaaaagatattgattattactggctcattcggtaacggtcatatgcaagttacacag 72
          |||||||   |||| ||| | ||||     | ||| || || ||| |||| ||  |  |
Db     10 aataaaagagtattaattttgactgcaaattacggaaatggacatgtgcaggtagccaaa 69

Qy     73 agtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattgagcacgat 132
          |   |  | | |||  |        |     |  | ||| |||  ||   |       ||
Db     70 acactttatgaacaatgtgtacggctcggctttcagcatgtaacagtttct------aat 123

Qy    133 ttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtatatcaatagc 192
          || |     ||  |  |||| ||| |  |     |   | || |    ||| | || |||
Db    124 ttgtaccaagagtcaaatccgattgtttcagaggtaactcaatacctttatttaaaaagc 183

Qy    193 tttaaatattttagaaatatgtacaaagggttttatt---------acagccgcccagat 243
          ||   |      | | |  | ||      ||||||||         |||    |    ||
Db    184 ttctcaatcgggaaacagttttatcgtttgttttattacggagttgacaaaatctataat 243

Qy    244 aaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgata 303
          ||||   |      | |||||    |    ||   ||| |  ||     |  |  |
Db    244 aaacgtaaattcaatatttacttttaaaatgggtaataaaagattgggcgaacttgtcgat 303

Qy    304 aaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaact 363
          || |  |||| ||| | ||  |    || ||||| |     ||  || ||| |   |
Db    304 gaacatcagcccgatattattattaatacatttccgatgatcgtcgtgccggaatacaga 363

Qy    364 gagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataaa 423
          | |  |   |     |||||       |  || ||||| || | | |  | ||||||
Db    364 cgccgaactggaagagtcattcctaccttcaacgttatgactgatttttgtcttcataaa 423

Qy    424 aactggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagac 483
          |  ||| ||       |       | ||||||||||| ||| |  |   |||  || |
Db    424 atttgggttcacgaaaacgtggataaatattatgtggcgacagattacgtgaaggaaaaa 483

Qy    484 ttcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataac 543
          |  | ||  | || | | ||||    |  || ||| | ||||| ||||| ||
Db    484 ctgctggagatcggcactcatccaagcaatgtaaaaatcacaggaattccaatcaggccg 543

Qy    544 aaatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccagat 603
          ||||||||    | ||   |      | |   |  | | |  |||| |    ||| |
Db    544 caatttgaagaatccatgcctgttggcccgatatataaaaagtacaatctttcaccaaac 603

Qy    604 aagcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatg 663
          ||  ||    || | ||    || |||||    |||||||  ||     |  |   | ||
Db    604 aaaaaagtgcttctgatcatggcaggtgctcacggtgtattaaagaacgtaaaagagctg 663

Qy    664 attactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagc 723
              || |   | |  ||   ||   | |  |||||||||  |  | || || || |
Db    664 tgc---gaaaaccttgtcaaggatgaccaagtgcaagtagttgtcgtgtgcgggaaaaat 720

Qy    724 aaagagctaaagcgttctttaacagctaagtttaaattaacgagaatgtatttga---tt 780
          |  |  ||||   ||||| |  ||           ||   |     | || |   ||
Db    721 acggctttaaaagaatctttgagtgcgcttgaagcggaaaatggtgacaaattaaaagtt 780

Qy    781 ctaggttataccaaacacatgaatgaatggatggcatcaagtcaacttatgattacgaaa 840
          || || |||    | | |||  ||||     |       |   |    |||||||| ||
Db    781 ctgggctatgtggagcgcattgatgagctatttcggatcacagattgcatgattaccaag 840

Qy    841 cctggtggtatcacaataactgaaggtttcgcccgttgtattccaatgattttcctaaat 900
          || || || || ||  | || ||||     |||  | |  | ||  | ||| |    ||
Db    841 cccggcggcattactttgacagaagccacagccattggagtgcctgtcattctgtacaaa 900

Qy    901 cctgcacctggtcaagagcttgaaaatgccttttactttgaaga 944
          || |  ||||| || ||    ||||||||    | |||||||||
Db    901 cccgtgcctggccaggaaaaagaaaatgcaaacttctttgaaga 944
```

Search completed: June 29, 2001, 11:49:52
Job time: 5214 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 10:55:47 ; Search time 1964.27 Seconds
(without alignments)
4692.090 Million cell updates/sec

Title: US-09-668-788-3
Perfect score: 975
Sequence: 1 atggttactcaaaataaaaa.....gtaaaacgctgatactccag 975

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	64.2	6.6	1101	219	CNS0039G	AL063921 Drosophil
2	61	6.3	976	221	CNS04E5M	AL286627 Tetraodon
3	60.4	6.2	1101	219	CNS0039G	AL063921 Drosophil
4	59.4	6.1	1101	219	CNS00KB5	AL077453 Drosophil
5	59	6.1	1101	219	CNS00EVL	AL069706 Drosophil
c 6	58	5.9	1101	219	CNS00EVL	AL069706 Drosophil
7	57	5.8	1101	219	CNS003BD	AL064091 Drosophil
8	56.8	5.8	1043	219	CNS0145P	AL103735 Drosophil
9	56.8	5.8	1092	220	CNS020K7	AL175696 Tetraodon
c 10	55.6	5.7	1101	219	CNS016LI	AL106896 Drosophil
c 11	55.4	5.7	548	120	AW787593	AW787593 945011D04
c 12	55.4	5.7	595	120	AW787592	AW787592 945011D04
c 13	55.4	5.7	610	104	AI979696	AI979696 614045D09
c 14	55.2	5.7	1101	219	CNS001FB	AL060732 Drosophil
c 15	54.8	5.6	1092	220	CNS020K7	AL175696 Tetraodon
c 16	54.2	5.6	876	219	CNS009G1	AL053529 Drosophil
17	53.4	5.5	1201	219	CNS015WP	AL106003 Drosophil
18	53.2	5.5	964	219	CNS006N9	AL065781 Drosophil
c 19	52.2	5.4	256	175	BG265588	BG265588 1000029G0
c 20	52.2	5.4	843	219	CNS00CS1	AL059666 Drosophil
21	52	5.3	987	219	CNS014PQ	AL104456 Drosophil
c 22	52	5.3	987	219	CNS014PQ	AL104456 Drosophil
23	51.8	5.3	1146	220	CNS021G2	AL176843 Tetraodon
c 24	51.4	5.3	966	219	CNS006R7	AL065822 Drosophil
c 25	51.4	5.3	970	219	CNS0182E	AL108800 Drosophil
c 26	51.2	5.3	550	31	AV542008	AV542008 AV542008
27	51.2	5.3	843	219	CNS00CS1	AL059666 Drosophil
28	51.2	5.3	1101	219	CNS00BO1	AL057419 Drosophil
c 29	50.8	5.2	935	219	CNS00ZX1	AL098239 Drosophil
30	50.8	5.2	945	221	CNS04D0K	AL285149 Tetraodon
c 31	50.8	5.2	1101	219	CNS0175J	AL107617 Drosophil
32	50.6	5.2	979	219	CNS0161W	AL106190 Drosophil
33	50.4	5.2	1200	219	CNS016CO	AL106578 Drosophil
34	50	5.1	928	219	CNS00DKY	AL071865 Drosophil
35	50	5.1	1190	220	CNS02ON7	AL206908 Tetraodon
c 36	49.8	5.1	569	16	AI108263	AI108263 GH07057.5
37	49.8	5.1	890	247	AZ670895	AZ670895 ENTLE93TF
c 38	49.8	5.1	1101	219	CNS000B8	AL063632 Drosophil
39	49.6	5.1	905	219	CNS00KHX	AL077798 Drosophil
40	49.6	5.1	1001	219	CNS0155H	AL105023 Drosophil
c 41	49.4	5.1	925	219	CNS006A9	AL064000 Drosophil
c 42	49.4	5.1	961	219	CNS008HI	AL051882 Drosophil
c 43	49.4	5.1	1200	219	CNS016CI	AL106572 Drosophil
44	49.4	5.1	1201	219	CNS0167M	AL106396 Drosophil
45	49.2	5.0	914	219	CNS00ZJY	AL097768 Drosophil

ALIGNMENTS

RESULT 1
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR08K10"
/note="end : TET3"
BASE COUNT 201 a 64 c 131 g 202 t 503 others
ORIGIN

Query Match 6.6%; Score 64.2; DB 219; Length 1101;
Best Local Similarity 16.6%; Pred. No. 0.00028;
Matches 99; Conservative 263; Mismatches 234; Indels 1; Gaps 1;

```
Qy    177 atggtatatcaatagctttaaatattttagaaatatgtacaaagggttttattacagccg 236
          ::::|:::: ::|   |:: :: ::  :||:|: :  :: |      |:::::: :   :
Db   1091 WDRDTRKDDWDWTKWWTWWKDRADDRRWAGDADRWAWDDGAGTWWTATWWWWWWWWATWD 1032

Qy    237 cccagataaactagacaaatgtttttacaaatactatggacttaataagttaattaattt 296
             ::::::|  |:   |:| : | :: ::| :    |::  :|||:|:   : :   :
Db   1031 TWWDKWWWWATAAKTDTAWTWWRTAWRADWAGRDRGAGKRDRDAATDADGAGRRDGGRKR 972

Qy    297 attgataaaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggt 356
           ::::  : :::   |   :| : | :||:::::: :::  :: |  :: ::::  |::
Db    971 KDKKDRKDGDDDKKGGKKKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDK 912

Qy    357 actaactgagcaatttaacattaatattccagttgctacagtgatgacagactatcgctt 416
          :  :   ::|: : : | : :::: ::::  |: |  :| : : ||:: :    :    |:
Db    911 DDDGKGKDADDDTDGTKDDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTW 852

Qy    417 acataaaaactggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaa 476
          : | :::| :: : :| :  : : : |: |::::: :: |::| |: ::::::|: :::
Db    851 DAAADDWWADDRWDAWAWKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRD 792

Qy    477 acaagacttcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctat 536
          | ::::   :|     | ::|::::::  |: :: |: |:::::: ::: |   ::
Db    791 AADDRDDAATWTTWTTTTTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWR 732

Qy    537 tgataacaaatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttaga 596
          :::: :::|  ::|    || ::: : ::| :  :|::   :: :: | :  :|
Db    731 RTWKRRWKRRDTRWDDADADDTARDDRRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDA 672

Qy    597 tccagataagcaaactattttaatgtcagctggtgcatttggtgtatctaaaggttttga 656
          :  : |::   : : ::::: :: :: |:  ::|  :  : :: |:: :::|: :::|:|
Db    671 WWADAAWWTTTDTDTDDWDKRDRRRKGARRRRRTTARAAW-DWWTWKAWDWAKWDWKTRA 613

Qy    657 cacgatgattactgatatattagcgaaaagtgcaaatgcacaagtagttatgatttgtgg 716
          : ::: |:|: : ::|::::|: :|::|: : |:| : : :: | |:| |  :||   :
Db    612 DRWDRWAADTWTDARKADRDWAKARAWRARRDRARAARADRRWTTKGKTTTATWTTWAAR 553

Qy    717 taagagcaaagagctaaagcgttctttaacagctaagtttaaattaacgagaatgta 773
           |: | |:|   |:     || |||      | : || |::: |:  | :|  ||
Db    552 AAWWAWWAWATTTATWTTTWTTTWTTTTTTTTTTTTWTTTAAWWWAAWWTATWAAWTA 496
```

```
RESULT   2
CNS04E5M
LOCUS       CNS04E5M      976 bp    DNA             GSS       21-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            103P02 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL286627
VERSION     AL286627.1  GI:8025084
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 976)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 976)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 976)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .976
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="103P02"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG103DH01LP1~end : T7"
BASE COUNT      243 a    151 c    137 g    327 t    118 others
ORIGIN
```

```
  Query Match             6.3%;  Score 61;  DB 221;  Length 976;
  Best Local Similarity   42.4%;  Pred. No. 0.0014;
  Matches  133;  Conservative   26;  Mismatches  155;  Indels    0;  Gaps    0;

Qy      23 tattgattattactggctcattcggtaacggtcatatgcaagttacacagagtatcgtta 82
           |:|: :|| |    |   | :||   | |   | |     :| |||   | | |    |::
Db     655 TWTWTWTTTTNTWTWWDTWWTTTATTTATTWTTAATNNAWAWTTAATAAAATTTAWNTWW 714

Qy      83 atcaacttaatgatatgaatctagaccatttaagcgtcattgagcacgatttatttatgg 142
           ||  | ||:|| |||| ||| :| |   |: :|   | :|| :  |  :|:| :|:||
Db     715 ATATAATTWATAATATTAATTWAWATWTTWAWATTTTTWTTTWAWATTWTWTTWTWATNA 774

Qy     143 aagctcatccaattttgacttctatttgtaaaaaatggtatatcaatagctttaaatatt 202
           |  | |:  :  |||    |   |   :|  ||||   | || |||   |||| | |||
Db     775 TAWWTAAWTWWTATTTTTTATTATTAAAWATTAAATAANAAATWAATTTTTTTATAAATT 834

Qy     203 ttagaaatatgtacaaagggttttattacagccgcccagataaactagacaaatgttttt 262
           |||  | :|: :   :|   |||||||           | ||| || | || : ||||
Db     835 TTATTATWAWWWTWTWAWATTTTTATTTATWWTWTNTTTANAAAWTATATAATWATTTTN 894

Qy     263 acaaatactatggacttaataagttaattaatttattgataaaagaaaagccagatttaa 322
           :||   ||      | ||| | | |||  | |  |  |       |    | ||| ||
Db     895 TTWAAATTTAATTTTTAAATTAATAAATATAATATTATTTTTTTTTTATTAAAAATTAAA 954
```

```
Qy     323 tattattaacgttt 336
           | | || |  |||
Db     955 AAATTTTTAAATTT 968
```

```
RESULT   3
CNS0039G
LOCUS       CNS0039G     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR08K10"
                     /note="end : TET3"
BASE COUNT      201 a     64 c    131 g    202 t    503 others
ORIGIN
```

```
  Query Match             6.2%;  Score 60.4;  DB 219;  Length 1101;
  Best Local Similarity   19.1%;  Pred. No. 0.0019;
  Matches  124;  Conservative  239;  Mismatches  285;  Indels    0;  Gaps    0;

Qy     194 ttaaatattttagaaatatgtacaaagggttttattacagccgcccagataaactagaca 253
           :::::| ||||  |:| |:  |  ||   :  :|: | |       | |:||| :|   :
Db     403 WWWWWTTTTTTTTTAWAAAWAAAATAATTWWAAWAWAAAAAATTWWAAAAWAAAAWAWTAW 462

Qy     254 aatgtttttacaaatactatggacttaataagttaattaatttattgataaaagaaaagc 313
           : |  :|: | ||| |  |   : ||  | : |||:||:|| ::|| ::  || :|||
Db     463 WTTTAWTWAAAAAAAAAAAAATTWTTTTTTTWTTTAWTTWATAWWTTWWWTTAAAWAAAAA 522

Qy     314 cagatttaatattattaacgtttcctacaccagttatgtcggtactaactgagcaattta 373
           | |  :|| :  |:  |  |:|  |:   :  : |: |     : |:::      |::
Db     523 AAAAAAWAAAWAAAWATAAATWTWWTWWTTYTTWAAWATAAAMCMAAWYYHTYTTYTYHY 582

Qy     374 acattaatattccagttgctacagtgatgacagactatcgcttacataaaaactggatta 433
             :|    :::: :|  :|:  : :: :   : ::   ::|: |:::: :|   ::
Db     583 YTYWTYTMTWHYHTMYTHAWAHTTWYHWYHTYAMWHWMTWHWTMWAWWHWTTYTAAYYYY 642

Qy     434 cgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagacttcatagacg 493
           : |: : :::::::| |:|  :: |   : :| ::   ::   :: | :|   :|  :|
```

```
Db    643 YTCMYYYHYMHWHHAHAHAAAWWTTHTWWTHAYHWATYHYYYYMYCAMMCMCTHTCHHCY 702

Qy    494 taggtattgatccttcaacagttaaagtgacaggtattcctattgataacaaatttgaaa 553
          :   : |  :|: ::  |:   :  :      :  || |:::::: :::  : |:: :|:
Db    703 YYYHHYTAHHTHTHHWYAHYYMWYYMWAYYWMYCTACTYHYHHHHHYHWAYHTTWYAWAH 762

Qy    554 cgcctattaatcaaaagcagtggttaatagacaacaacttagatccagataagcaaacta 613
          :::: :   |||:  : |  ::::|   :::  ::::   |||   | :  :  :||
Db    763 AMWMWHHAHYAAAAAWAAWATTHHYHHTTHYMHHTYMYHYYMYTCCYMCTYHCWHYYHTA 822

Qy    614 ttttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatgattactgata 673
          :| : :| : :  :  : ::|::  | |:: :::   : :: ::|  :| ||:    :|:
Db    823 YTCWTWTHHWMWTWTHWYHHTWWHHTTTHWAWWHTHTWCWWWWHATTWTWATHCWACMTM 882

Qy    674 tattagcgaaaagtgcaaatgcacaagtagttatgatttgtggtaagagcaaagagctaa 733
          ::::: : ::::   :|::: ::|:  :: :    :::| :  | : : :::: : ::::
Db    883 HWHHWMHMHHHHMACHAHHHTHMCMCHHHMHCTCHHHHTMYHMTCHWMWHMHWHWHHWWM 942

Qy    734 agcgttctttaacagctaagtttaaattaacgagaatgtatttgattctaggttatacca 793
          |   ||:|| :::  : ::  :  :: :: :   : : | :|:  |:::   |  : | :
Db    943 ATWMTTMTTMMMMCCMMHHHCHMYHMMHMYMYCCHYYCTCHTHATTHYHYMCTCYHYCTW 1002

Qy    794 aacacatgaatgaatggatggcatcaagtcaacttatgattacgaaac 841
          : :: |: :|: |:        :::::  ::| :::: :::|  ::||
Db   1003 HTYWTAYWWAWTAHAMTTATWWWWMHWWAHWATWWWWWWWWATAWWAC 1050


RESULT  4
CNS00KB5
LOCUS       CNS00KB5     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR39P05 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL077453
VERSION     AL077453.1  GI:4956930
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR39P05"
                     /note="end : T7"
BASE COUNT      540 a      0 c    139 g    252 t    170 others
ORIGIN
```

```
  Query Match             6.1%;  Score 59.4;  DB 219;  Length 1101;
  Best Local Similarity   38.7%;  Pred. No. 0.0032;
  Matches  232;  Conservative   28;  Mismatches  340;  Indels    0;  Gaps    0;

Qy    172 aaaaaatggtatatcaatagctttaaatattttagaaatatgtacaaagggttttattac 231
          ||||||      || ||||   | ||| | ||   |||||| || |:|   |   | ||
Db    143 AAAAAAAAAATAATAAATAATATAAAAAAATTATAAAATATTTAAAWATWATAAAAATAA 202

Qy    232 agccgcccagataaactagacaaatgtttttacaaatactatggacttaataagttaatt 291
          |       | || |:  | | |||:  ||  | ||| : ||   |  :|| :|  :||
Db    203 ATWWATTAAAATTAWWAAAAWAAAWAATTAAAAAAAAAWATAAAAAAAWAAAWATAWAAAA 262

Qy    292 aatttattgataaaagaaaagccagatttaatattattaacgtttcctacaccagttatg 351
          ||  ||   | |||| ||||   | |   ||||  |  ||   |    | |  |    |
Db    263 AAAATAAAAAAAAAAAAAAAAAATAAAAAAAATAAAAAAAAATATAAAAAAAAAAAAAAA 322

Qy    352 tcggtactaactgagcaatttaacattaatattccagttgctacagtgatgacagactat 411
          |       |  ||   || | | |||    | ||   | | |  |       | |
Db    323 AAAANANNNNNNAAATAAAAAAANAATTATANANAATTTATAATATTTTTTTATTTTTTT 382

Qy    412 cgcttacataaaaactggattacgccgtattcaacaagatattatgtggcaacaaaagaa 471
          |||  ||   ||    || |      |   || || |||  |      || |||| ||
Db    383 ATTTTATATTTTAAAAAAATAAAAATAAAAAAAAAAAAAATAAAAAAAAATAAAAAAAAA 442

Qy    472 acgaaacaagacttcatagacgtaggtattgatccttcaacagttaaagtgacaggtatt 531
          |   || || |   | | |   |   |   |       || |   ||  | |   |
Db    443 AWATAAAAAAAAAAAAAAAAAATNAAANAAANAANNNNAAAAAANAAATATAATTTAAAAA 502

Qy    532 cctattgataacaaatttgaaacgcctattaatcaaaagcagtggttaatagacaacaac 591
          ||  |    ||   |||      |  ||  |||| |      || | | || ||
Db    503 AAATTTTTTTTNTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAATAANAAA 562

Qy    592 ttagatccagataagcaaactattttaatgtcagctggtgcatttggtgtatctaaaggt 651
          |   |   ||  |||   :  | |:|  |: :     ||||: ::|    |||  |
Db    563 AAAANANAAANAAAAAAAAANAGKGATGAKGRKARRKWTAAAATTTKTKRTKGKGAAAATT 622

Qy    652 tttgacacgatgattactgatatattagcgaaaagtgcaaatgcacaagtagttatgatt 711
          |: |      :  |:  |: |   :  |      : | |: |          || | ||
Db    623 TKGGGTTGTTWTTTWTATKGTGGTWGGGKWTTTKKGGKADTTAGKTKGTGTTTTTTTATA 682

Qy    712 tgtggtaagagcaaagagctaaagcgttctttaacagctaagtttaaattaacgagaatg 771
          |: |||  | | :  | |  || : | :   |    |  | | |   :||:| :  ||||
Db    683 TKGGGTGGGGGGRTKGGGGRAATKGGGKKGGTGTATGTGAKGGTGGTRTTRAGKTWAATG 742


RESULT  5
CNS00EVL
LOCUS       CNS00EVL     1101 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
```

NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR29B23"
/note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN

Query Match 6.1%; Score 59; DB 219; Length 1101;
Best Local Similarity 34.1%; Pred. No. 0.0039;
Matches 126; Conservative 75; Mismatches 165; Indels 3; Gaps 1;

```
Qy      11 aaaataaaaagatattgattattactggctcattcggtaacggtcatatgcaagttacac 70
           ||||:||:|| : ::  | |||:: |   | :||   :::   : :::|  |: | : |
Db     736 AAAAWAAWAATWAWWATAATATWWATATATAWTTWTAWWWATWWAWWWTATAWATAWAAT 795

Qy      71 agagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattgagcacg 130
           | | :|:   |||  |: | :|| | |: :|: || :  : | |     |:  | |
Db     796 AWAAWAWAWATAAATAWATAWATWAAAWAWAWATAWWATWATATAWWAATAWAWAAAAAA 855

Qy     131 atttatttatggaagctcatccaatttgacttctatttgtaaaaaatggtatatcaata 190
           :| || |:   ||    |   ||   | : :| ::|:| | ::|::   || |: :|||
Db     856 TWTAATATW---AATWATAAWAAAAAATAWAWTTWWTWTTTTWWAWWATATAAAWAWATA 912

Qy     191 gctttaaatattttagaaatatgtacaaagggttttattacagccgcccagataaactag 250
           :||| |    | ||  |: ::  :|   |:|||||| |       :  |:|  ::
Db     913 WAAAWAAAAAAAAAAAAATAAWAWWWTWATATTWTTATTAAAWTWTATWWATTWATTWWA 972

Qy     251 acaaatgtttttacaaatactatggacttaataagttaattaatttattgataaaagaaa 310
           : : || :||:|| : |||   :  | | :|||  || |||||::|||   | |:| |:
Db     973 WTWTATAWTTWTATWTATATATWWTAWTAWATATATTTATTAAWWTATATTTTAWAAAWT 1032

Qy     311 agccagatttaatattattaacgtttcctacaccagttatgtcggtactaactgagcaat 370
           |     : ||:::| : ||    ::  |  :  | |||       ::|   |  ||:
Db    1033 AATATATWATAWWWTAWATATAWAWWAATTAWTTATATATWTAAWATAWWAAAAAWAAAW 1092

Qy     371 ttaacatta 379
           :|  :: |
Db    1093 AWATAWWAA 1101
```

RESULT 6
CNS00EVL/c
LOCUS CNS00EVL 1101 bp DNA GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR29B23"
/note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others
ORIGIN

Query Match 5.9%; Score 58; DB 219; Length 1101;
Best Local Similarity 31.5%; Pred. No. 0.0065;
Matches 185; Conservative 106; Mismatches 293; Indels 4; Gaps 1;

```
Qy      65 ttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattg 124
           | | | | | : |  | |||  | :||:| ::||  || :| |  |:|  |   | :
Db    1026 TAAAATATAWWTTAATAAATATATWTAWTAWWATATATAWATAWAAWTATAWAWTWWAAT 967

Qy     125 agcacgatttatttatggaagctcatccaatttgacttctatttgtaaaaaatggtatata 184
           :    || :|:||    ||    ||  :::|:| | || | ||| |  :   :  |:|:
Db     966 WAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTTTTTTWTTTWTATWTW 907

Qy     185 tcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagata 244
           |  |||  :|::|| |: ::| : :||| |   :    : ||: ||          :|:
Db     906 TTTATATWWTWWAAAAWAWWAAWTWTATTTTTTWTTATWATTWATATTAWATTTTTTWTW 847

Qy     245 aactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgataa 304
           | |:    | : |::|    | :  |:|  : |: ||:  || ||: :|:| :|  |:
Db     846 TATTWWTATATWATWWT----ATWTWTWTTTWATWTATWTATTTATWTWTWTTWTATTWT 791

Qy     305 aagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactg 364
           |   | |     : |:::||: |:||   |:  ||    : |:|| :   |  ||  :
Db     790 ATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWTTWTTTTATTWT 731

Qy     365 agcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataaaa 424
           ||:|||    |:|| :|  | :|  |   |   |   : | | :  |||  |||
Db     730 TTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTTWTATTTWAAATTATTTAATT 671

Qy     425 actggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagact 484
           | :  ||:|     |    ||  |  ||   | |   ||       :|  ||| |   |
Db     670 AWWWAATWAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTATWAAAAAATATTTTT 611

Qy     485 tcatagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataaca 544
           | |||    |:    ||     || |   ::  :   |    || |    |:  | | |
Db     610 TTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWTATTWWTTTAAA 551

Qy     545 aatttgaaacgcctattaatcaaaagcagtggttaatagacaacaacttagatccagata 604
           |:        ||:||:  : :|  :     ::::: | || :: :|:  |  :  |:
Db     550 TTATWTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWTWTTTTTWATTW 491

Qy     605 agcaaactattttaatgtcagctggtgcatttggtgtatctaaaggtt 652
           |  |:  |:|||::|: | :  ::::: : :||    || |:  ::::
Db     490 ATAAWTTTWTTTWWAWATTWWKKKKKKAWADTGAGARATKTDKKKKKK 443
```

RESULT 7
CNS003BD
LOCUS CNS003BD 1101 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #

```
            BACR08K08 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL064091
VERSION     AL064091.1  GI:4941847
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR08K08"
                      /note="end : TET3"
BASE COUNT      395 a    120 c    103 g    334 t    149 others
ORIGIN
```

```
  Query Match             5.8%;  Score 57;  DB 219;  Length 1101;
  Best Local Similarity   38.1%;  Pred. No. 0.011;
  Matches  177;  Conservative   50;  Mismatches  233;  Indels    4;  Gaps    1;

Qy    165 tatttgtaaaaaatggtatatcaatagctttaaatattttagaaatatgtacaaagggtt 224
          | ||| |      |  | | |   ||  ||| :    |||:  ||| | || | |   ||
Db    605 TTTTTTTTTTTTTTTTTTTTTTTTAWTTTTWTATNTTTWTTAATTTTTAAATAAWTTT 664

Qy    225 ttattacagccgcccagataaactagacaaatgtttttacaaatactatggacttaataa 284
          |  |||          | | |||  | | |||      |   ||| |  |   |   |:|:
Db    665 TATTTAWNAAAWATTAAAAAAAAAAAAAAAAAAAAAAATATAAAAAAAAAATTATAAAWTWT 724

Qy    285 gttaattaatttattgataaaagaaaagccagatttaatattattaacgtttcctacacc 344
             ||  |:|: || : || | |  |    | |: ||:| ||||| |  ||    | |
Db    725 AAATATATAWTWTTTAWAAATATATTAAAAAAAWATAWTTTTATTTATATTAAAAATATW 784

Qy    345 agttatgtcggtactaactgagcaatttaacattaatattccagttgctacagtgatgac 404
          | :  |       : : | :   :|||||  ||:|||:|  | |: :   |   ||
Db    785 ATWATTTATATATNWNWATAWTTTWTTTAAATTTWATAWTTAAATWTMATTAAAWATTNT 844

Qy    405 agactatcgcttacataaaaactggattacgccgtattcaacaagatattatgtggcaac 464
          | | ||    : | | || ||    | ||     |:|:  | :   :|     :   |:
Db    845 AWAATAAAAAWAAAAAAATAATAAAAATAATWTATWTWATAWWTTTWAAATAAWTAAAWT 904

Qy    465 aaaagaaacgaaacaagacttcatagacgtaggtattgatccttcaacagttaaagtgac 524
          ||:| |||  ||  || |    |  |  ||  |:|| :|:  : ||    | |: : :
Db    905 AAWAAAAAAAAATTAAWATAAAAATWAAWTAWTTWTTTMTYTAWAAAWYWATTAWAWAWA 964

Qy    525 aggtattcctattgata----acaaatttgaaacgcctattaatcaaaagcagtggttaa 580
          |  | ||  |||:  :|    |  |::   |||    ||  | : ||    : :  : ||
Db    965 ATATTTTTATATWATWATTATAATAWWAAAAAAAAAATAAAATWAAATNWAWAWWAYAAA 1024
```

```
Qy    581 tagacaacaacttagatccagataagcaaactattttaatgtca 624
          :| ||:| || :|: || :| |::|  |||: |:|  | | | |
Db   1025 WAYACWAAAAAWTMTATWYATAWMAAAAAAMAAHTANATTATWA 1068


RESULT   8
CNS0145P
LOCUS       CNS0145P     1043 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN11G11 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL103735
VERSION     AL103735.1  GI:5615346
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1043)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES              Location/Qualifiers
     source           1. .1043
                      /organism="Drosophila melanogaster"
                      /plasmid="pBeloBAC11"
                      /db_xref="taxon:7227"
                      /clone_lib="DrosBAC"
                      /clone="BACN11G11"
                      /note="end : T7"
BASE COUNT      277 a     96 c    121 g    382 t    167 others
ORIGIN
```

```
  Query Match             5.8%;  Score 56.8;  DB 219;  Length 1043;
  Best Local Similarity   34.2%;  Pred. No. 0.012;
  Matches  143;  Conservative   66;  Mismatches  209;  Indels    0;  Gaps    0;

Qy     75 tatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattgagcacgattt 134
          | |  |: ||  |  ||:: ||:|  ||  | |  :| ||     |  ||    |  |  :
Db    551 TTTTTTWTATATAWATAWWTATWTATATAAAAATAMTATATATATTTTTATATATWAAAW 610

Qy    135 atttatggaagctcatccaattttgacttctatttgtaaaaaatggtatatcaatagctt 194
          | |  :   |  |  |   :::|: : || |:|::  || | ||  | |:|  :|:  ||
Db    611 AATATWTATATWTTTTTTTWWWTWTWTTTATWTWWWAAATATATWTTTTWTWTWTTTTT 670

Qy    195 taaatattttagaaatatgtacaaagggttttattacagccgcccagataaactagacaa 254
          |: :||| ||| :: : : :| |||   ||||  |: |       | :|| : |  |
Db    671 TWTWTATATTATWWTWTWTWATAAATATTTTTTATWTATATTTATATWTATWTTTWAATT 730

Qy    255 atgtttttacaaatactatggacttaataagttaattaatttattgataaaagaaaagcc 314
             | ||  | ||| | |    ||  |   ||  ||  ||| ||  ||||| |:||
Db    731 TATATATTTWANATANTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAAWAAAAW 790

Qy    315 agatttaatattattaacgtttcctacaccagttatgtcggtactaactgagcaatttaa 374
          : :::::: |:::::::  :              ||:: |:  :   :|   |  :|: ||:
Db    791 WWWWWWWWAAWWWWWWWWAWAAWAATTTTTTTTTHHTTYCCYYTAWATATATWWAWATAW 850

Qy    375 cattaatattccagttgctacagtgatgacagactatcgcttacataaaaactggattac 434
          : :|| |    |       | |   |  | | |  |      |:|:| | ||   |
Db    851 AWAWAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAMAHAYATACMAAAAMCW 910
```

```
Qy    435 gccgtattcaacaagatattatgtggcagcaaaagaaacgaaacaagacttcatagac 492
          :  :|  :| | : |||    ||  |||| |||  ||:::: ::   | | ||
Db    911 AATHAWAAMAYAHAAAAWATATAAAAAAAAAAAAAAAAAAAAMMMMMMMMAAAAAAAC 968


RESULT  9
CNS020K7
LOCUS       CNS020K7     1092 bp    DNA             GSS       12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            222L11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL175696
VERSION     AL175696.1  GI:7813753
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1092)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1092)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="222L11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG222CF06LP1~end : T7"
BASE COUNT      383 a    169 c    165 g    262 t    113 others
ORIGIN


  Query Match             5.8%;  Score 56.8;  DB 220;  Length 1092;
  Best Local Similarity   36.5%;  Pred. No. 0.012;
  Matches  178;  Conservative   62;  Mismatches  246;  Indels    2;  Gaps    1;

Qy     68 cacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattgagc 127
          |::| | : |  : ||:    |:  | |::: ||   | |  |  ::    |  |  |
Db    599 CMMAAAHWTTTTWAAAWTTTATWTTTAAWWWWAAAAAAAAAAAAAAWWTTTTTTTTAWAAA 658

Qy    128 acgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtatatca 187
          |  |:  :::::        |   ||||||   || | ||| |   |||    | |  |
Db    659 AAAAWAAWWWWWWWWTTAAAAAAAAAATTTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAA 718

Qy    188 atagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagataaac 247
          | |  |: | : | :: | |||:|   | ::|   :||| ||: :       | |:|||
Db    719 AAAAWTWAATWAAAWWAAAAAAWAAWAAWWWAWTTWTTTTTTTWAWWWWWWWAAAAWAAAW 778

Qy    248 tagacaaatgtttttacaaatactatggacttaataagttaattaatttattgataaaag 307
          :: :  |:| || :::  :|:: |:|    | ||:||   |:|| |:|||:| | |||:
Db    779 WWTWTTAWTTTTAWWWATWAWWTTWTATTTTAAAWAAAANAWTTTAWTTAWTAAAAAAWW 838
```

```
Qy    308 aaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactgagc 367
          |||       || || : : |||:   ||   | :  |   |      ||  || : |
Db    839 AAATAATTWTTTAAAAWAWTTTAWAAATTWWAATWTAAAAAAAAAAATTAAAAAAWAAWT 898

Qy    368 aa--tttaacattaatattccagttgctacagtgatgacagactatcgcttacataaaaa 425
          ::  |||   :|:|:: :|  | :|  |    |  | :     |:|     | | |||||
Db    899 WWTTTTTTTTWTWAWWTWTAAATWTTTTTTTTTTTTTTWWTTTTWTAAWAAAAAAAAAAA 958

Qy    426 ctggattacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagactt 485
                |      ::| ||     | |||      || ||||  ||   |  || |  |
Db    959 AATATAAAWNAAAAWWTTAATTTATTTTTAAAAAAAAAAAAAAAATAAAATATAAAWAAAT 1018

Qy    486 catagacgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataacaa 545
          | | |  ||  |||| |      || |  |||| | | |  ||     |   | || ||
Db   1019 AAAATATTTATATATTAAAAAAAAAAAAAAATAAATTAAAAAWTAAAATAAAAAAAAAAAA 1078

Qy    546 atttgaaa 553
          |    |||
Db   1079 AAAAAAAA 1086


RESULT  10
CNS016LI/c
LOCUS       CNS016LI     1101 bp    DNA             GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106896
VERSION     AL106896.1  GI:5624374
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN16D22"
                     /note="end : T7"
BASE COUNT      203 a    220 c     84 g    158 t    436 others
ORIGIN


  Query Match             5.7%;  Score 55.6;  DB 219;  Length 1101;
  Best Local Similarity   31.6%;  Pred. No. 0.022;
  Matches  101;  Conservative   75;  Mismatches  144;  Indels    0;  Gaps    0;

Qy     23 tattgattattactggctcattcggtaacggtcatatgcaagttacacagagtatcgtta 82
          ||||   |||  :|   ::::::   |   :     :|:   | | : |  :   : |
Db   1101 TATTATATATATHTAYAWYWWYYTTMAYWWWWTYCTMTMAMTAMAAAWMAAAMWMAAWMA 1042

Qy     83 atcaacttaatgatatgaatctagaccatttaagcgtcattgagcacgatttatttatgg 142
          |: :| |||    |:: |||:||   ::    |     :|| |     ||:||||:||
Db   1041 AWAWATTTAYATTTHWTAATHTAWTWYWCAATAHWWAWWTTTATATWAATHTATTWATAH 982
```

```
Qy    143 aagctcatccaattttgacttctatttgtaaaaaatggtatatcaatagctttaaatatt 202
          :   :  :  :: |::  :|: |||:   :|||:|| :|| |: :|||  : :|:||| |
Db    981 WTATYWTWTWHWATWYWTMTYWTATWAWAHAAAWATASTAAAYWHATAATWAWAWATAAT 922

Qy    203 ttagaaatatgtacaaagggttttattacagccgcccagataaactagacaaatgttttt 262
          |:  : :||: |: :|:   | |:||:| :       : ::|:: || | :::: :| |:
Db    921 TWTWWTWTAWWTWTWAWTAATATWATWAAMTAATWWWWAWHAWWWTAYAWWWWWAWTATW 862

Qy    263 acaaatactatggacttaataagttaattaatttattgataaaagaaaagccagatttaa 322
          : | |:       | |: |    | : : || |   | :|:||| ||||    : |:  :|
Db    861 WTATAWTAATAWWATTWTAATTATAWTWAAAATATATAWTWAAAAAAAAWATHAAWAAWA 802

Qy    323 tattattaacgtttcctaca 342
           ||  : ||     ::|| |
Db    801 AATATWAAAAAAAAHMTAAA 782


RESULT 11
AW787593/c
LOCUS       AW787593      548 bp    mRNA            EST       16-MAY-2000
DEFINITION  945011D04.X2 945 - Mixed adult tissues from Walbot lab, same as 707
            (SK) Zea mays cDNA, mRNA sequence.
ACCESSION   AW787593
VERSION     AW787593.1  GI:7844371
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 945011  row: D  column: 04.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="945 - Mixed adult tissues from Walbot lab,
                     same as 707 (SK)"
                     /tissue_type="tassel, kernal, silk, husk, root, leaf"
                     /dev_stage="fully-grown"
                     /lab_host="DH10B"
                     /note="Organ: tassel, kernal, silk, husk, root, leaf;
                     Vector: pGAD10; Site_1: EcoRI; cDNA library from fully
                     differentiated maize tissues from an active Mutator plant.
                     Tissue ratio is 4:2:1:1:1:1 (tassel, kernel, silk, husk,
                     root, leaf). Unidirectionally cloned.  New library number
                     given to library 707  for additional sequencing."
BASE COUNT      163 a    131 c     97 g    157 t
ORIGIN


  Query Match             5.7%;  Score 55.4;  DB 120;  Length 548;
  Best Local Similarity   57.1%;  Pred. No. 0.023;
  Matches  101;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    783 aggttataccaaacacatgaatgaatggatggcatcaagtcaacttatgattacgaaacc 842
          ||| | |   | | | ||| | ||||| ||||   |  || |  ||| ||||| ||  |
Db    519 AGGCTTTGTTACAAAGATGGAAGAATGTATGGGTGCTTGTGATTGTATCATTACAAAGGC 460

Qy    843 tggtggtatcacaataactgaaggtttcgcccgttgtattccaatgattttcctaaatcc 902
          ||   |   ||||| | || |   |   |||| |  | ||||| ||| |     ||
Db    459 AGGACCTGGTACAATTGCAGAGGCAATGATCCGTGGCTTACCAATTATTCTAAATGATTA 400

Qy    903 tgcacctggtcaagagcttgaaaatgccttttactttgaagaaaaaggttttggtaa 959
          |    |||| || ||   || |||| |   ||| |||  ||||| || | ||| ||
Db    399 TATTGCTGGACAGGAAGCTGGCAATGTCCCCTACGTTGTTGAAAATGGATGTGGGAA 343


RESULT 12
AW787592/c
LOCUS       AW787592      595 bp    mRNA            EST       16-MAY-2000
DEFINITION  945011D04.X1 945 - Mixed adult tissues from Walbot lab, same as 707
            (SK) Zea mays cDNA, mRNA sequence.
ACCESSION   AW787592
VERSION     AW787592.1  GI:7844370
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 595)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 945011  row: D  column: 04.
FEATURES             Location/Qualifiers
     source          1. .595
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="945 - Mixed adult tissues from Walbot lab,
                     same as 707 (SK)"
                     /tissue_type="tassel, kernal, silk, husk, root, leaf"
                     /dev_stage="fully-grown"
                     /lab_host="DH10B"
                     /note="Organ: tassel, kernal, silk, husk, root, leaf;
                     Vector: pGAD10; Site_1: EcoRI; cDNA library from fully
                     differentiated maize tissues from an active Mutator plant.
                     Tissue ratio is 4:2:1:1:1:1 (tassel, kernel, silk, husk,
                     root, leaf). Unidirectionally cloned.  New library number
                     given to library 707  for additional sequencing."
BASE COUNT      176 a    142 c    104 g    173 t
ORIGIN


  Query Match             5.7%;  Score 55.4;  DB 120;  Length 595;
  Best Local Similarity   57.1%;  Pred. No. 0.023;
  Matches  101;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    783 aggttataccaaacacatgaatgaatggatggcatcaagtcaacttatgattacgaaacc 842
          ||| | |   | | | ||| | ||||| ||||   |  || |  ||| ||||| ||  |
Db    516 AGGCTTTGTTACAAAGATGGAAGAATGTATGGGTGCTTGTGATTGTATCATTACAAAGGC 457

Qy    843 tggtggtatcacaataactgaaggtttcgcccgttgtattccaatgattttcctaaatcc 902
          ||   |   ||||| | || |   |   |||| |  | ||||| ||| |     ||
Db    456 AGGACCTGGTACAATTGCAGAGGCAATGATCCGTGGCTTACCAATTATTCTAAATGATTA 397

Qy    903 tgcacctggtcaagagcttgaaaatgccttttactttgaagaaaaaggttttggtaa 959
          |    |||| || ||   || |||| |   ||| |||  ||||| || | ||| ||
Db    396 TATTGCTGGACAGGAAGCTGGCAATGTCCCCTACGTTGTTGAAAATGGATGTGGGAA 340


RESULT 13
AI979696/c
```

```
LOCUS       AI979696      610 bp    mRNA            EST       30-AUG-1999
DEFINITION  614045D09.x4 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AI979696
VERSION     AI979696.1  GI:5804726
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614045  row: D  column: 09.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      182 a    145 c    107 g    175 t      1 others
ORIGIN
```

```
  Query Match             5.7%;  Score 55.4;  DB 104;  Length 610;
  Best Local Similarity   57.1%;  Pred. No. 0.023;
  Matches  101;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    783 aggttataccaaacacatgaatgaatggatggcatcaagtcaacttatgattacgaaacc 842
          ||| | |   | | | ||| | ||||| ||||   |  || |   ||| ||||| ||  |
Db    523 AGGCTTTGTTACAAAGATGGAAGAATGTATGGGTGCTTGTGATTGTATCATTACAAAGGC 464

Qy    843 tggtggtatcacaataactgaaggtttcgcccgttgtattccaatgattttcctaaatcc 902
          ||   |   |||||  | || |   |  ||||  |  | |||||  ||| |     ||
Db    463 AGGACCTGGTACAATTGCAGAGGCAATGATCCGTGGCTTACCAATTATTCTAAATGATTA 404

Qy    903 tgcacctggtcaagagcttgaaaatgccttttactttgaagaaaaaggttttggtaa 959
          |    |||| || ||   ||  |||| |   ||| |||  ||||| || | ||| ||
Db    403 TATTGCTGGACAGGAAGCTGGCAATGTCCCCTACGTTGTTGAAAATGGATGTGGGAA 347
```

```
RESULT  14
CNS001FB/c
LOCUS       CNS001FB     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR04A23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL060732
VERSION     AL060732.1  GI:4939397
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR04A23"
                     /note="end : TET3"
BASE COUNT      288 a    110 c    103 g    491 t    109 others
ORIGIN
```

```
  Query Match             5.7%;  Score 55.2;  DB 219;  Length 1101;
  Best Local Similarity   37.5%;  Pred. No. 0.027;
  Matches  158;  Conservative   51;  Mismatches  211;  Indels    1;  Gaps    1;

Qy      6 tactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatgcaagt 65
          ||   |:|  ||||:  |||   :||| : :   | | |     |      :||  |
Db   1100 TATATAWATAAAAAWWTTATAATWTATAWWWAAATWAATAAAATATWAAATWATAWATAA 1041

Qy     66 tacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattga 125
          || | | | |||   |:|| ||    |  ||:| ||  :| :  |:  ||   | | | |
Db   1040 TAAAAAAAATATAAATWATTAAAAATAAAATWTAAAAAWATWAAAWAAAAAATTWAATAA 981

Qy    126 gcacgatttatttatggaagctcatccaattttgacttctatttgtaaaa-aatggtata 184
            | || |    |   ||    |  ||||   | ::  |     ||:| |||  ||||
Db    980 TTAWTATATTNAAAAAAAAWAAAAAAAAATTAATAAWWAAAAAAAAAAAWATAATAATATA 921

Qy    185 tcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagata 244
          | ||::   |: |||:  ||: ||| ||    :||   :    | | |       | : |
Db    920 TAAAWWTAATWTAATWNATTWAAAAAATAATAWAATAWWAAATTAAWAAAAATAWATWAA 861

Qy    245 aactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgataa 304
          || :  | : ||  :||:| ||||:   :  :    |+  : ||  |||||  |||  :|||
Db    860 AAAWNTAWWTATAAWTTWAAAAATWTNTWAWWWAAAAATWWTTTTTTAATATATATWTAA 801

Qy    305 aagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactg 364
           : || :  . | |   ||||: |: :  :|:  :: | | : |     ||  ||
Db    800 TWTAATWTWWAAAAAAAATAWAAWAWTAWWTWAWWWNAWAAAWAAATAATATAWNAAWAA 741

Qy    365 agcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacataaaa 424
          |    : ::| ||:::|:::  |  |   | |   || :   | ||      | | ||||
Db    740 AAATTWAWWANATWWWTWWWTAAAATAAAAAAAAAATAWATAAATAAWAAAAAAAAAAAA 681

Qy    425 a 425
          |
Db    680 A 680
```

```
RESULT  15
CNS020K7/c
LOCUS       CNS020K7     1092 bp    DNA             GSS       12-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            222L11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
```

```
ACCESSION   AL175696
VERSION     AL175696.1  GI:7813753
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1092)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1092)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1092)
  AUTHORS   Genoscope.
  TITLE.    Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="222L11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG222CF06LP1~end : T7"
BASE COUNT      383 a    169 c    165 g    262 t    113 others
ORIGIN
```

```
  Query Match             5.6%;  Score 54.8;  DB 220;  Length 1092;
  Best Local Similarity   33.7%;  Pred. No. 0.033;
  Matches  128;  Conservative   64;  Mismatches  188;  Indels    0;  Gaps    0;

Qy       5 ttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatgcaag 64
           ||  | ||||||||   |::||   : |||     |  ||   | :   : | |   ||
Db     992 TTTTTAAAAATAAATTAAWWTTTTNWTTTATATTTTTTTTTTTTTTWTTAWAAAAAWWAAA 933

Qy      65 ttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattg 124
             | | | :||   :|:: :|   ||  | :: : | |     | ||     |  |
Db     932 AAAAAAAAAAWATTTAWAWWTWAWAAAAAAAAWWAWTTWTTTTTTAATTTTTTTTTTAW 873

Qy     125 agcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtata 184
           |  :  ||||:|  |:        |:     ||| : || | :|  : |||:|  | |:
Db     872 ATTWWAATTTWTAAAWTWTTTTAAAWAATTATTTWWTTTTTAWTAAWTAAAWTNTTTTW 813

Qy     185 tcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagata 244
           |  | |   : |::|: |::: ||| :|  : :::   :||| ::: :       | |:|
Db     812 TTTAAAATAWAAWWTWATWWWTAAAAWTAAWAWWWTTTWTTTTWWWWWWWTWAAAAAAWA 753

Qy     245 aactagacaaatgtttttacaaatactatggacttaataagttaattaatttattgataa 304
           |: ::   :  : |||||:    :| |:     || |  ||  ||||  |   | ||
Db     752 AWTWWWTTWTTWTTTTTTWWTTTWATTWAWTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAA 693

Qy     305 aagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactg 364
           || ||||      ||| | ::::::::  :|| |  :  |   |  :   |  |   |
Db     692 AAAAAAATTTTTTTTTTAAWWWWWWWWTTWTTTTTTTWTAAAAAAAAWWTTTTTTTTTTT 633

Qy     365 agcaatttaacattaatatt 384
             :::|||| : ||| :||
Db     632 TTWWWWTTAAAWATAAAWTT 613
```

```
Search completed: June 29, 2001, 10:55:58
Job time: 5021 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: June 29, 2001, 11:46:43 ; Search time 119.12 Seconds
(without alignments)
1516.360 Million cell updates/sec

Title: US-09-668-788-3
Perfect score: 975
Sequence: 1 atggttactcaaaataaaaa.....gtaaaacgctgatactccag 975

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	437	44.8	2600	3	US-08-988-251-1	Sequence 1, Appli
c	2	437	44.8	2600	4	US-09-386-048-1	Sequence 1, Appli
	3	46.4	4.8	8920	2	US-08-446-855A-1	Sequence 1, Appli
	4	46.4	4.8	8920	4	US-09-150-741-1	Sequence 1, Appli
c	5	44.8	4.6	7218	1	US-08-232-463-14	Sequence 14, Appl
	6	43.8	4.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	7	42	4.3	19124	2	US-08-487-826B-13	Sequence 13, Appl
	8	41	4.2	2058	2	US-08-749-391-1	Sequence 1, Appli
	9	41	4.2	2058	4	US-09-390-200-1	Sequence 1, Appli
c	10	40	4.1	8654	1	US-08-920-812-6	Sequence 6, Appli
c	11	40	4.1	8654	1	US-08-920-827-6	Sequence 6, Appli
c	12	40	4.1	8654	1	US-08-921-177-6	Sequence 6, Appli
c	13	40	4.1	8654	1	US-08-362-577C-6	Sequence 6, Appli
c	14	40	4.1	8654	2	US-08-920-828-6	Sequence 6, Appli
c	15	39.6	4.1	2022	2	US-08-505-486-96	Sequence 96, Appl
c	16	39.6	4.1	2022	3	US-08-801-028-96	Sequence 96, Appl
c	17	39.6	4.1	2022	3	US-09-340-154-96	Sequence 96, Appl
c	18	39.6	4.1	2022	5	PCT-US95-09338-96	Sequence 96, Appl
c	19	39.6	4.1	2022	5	PCT-US95-09339-96	Sequence 96, Appl
c	20	39.6	4.1	2127	2	US-08-505-486-95	Sequence 95, Appl
c	21	39.6	4.1	2127	3	US-08-801-028-95	Sequence 95, Appl
c	22	39.6	4.1	2127	3	US-09-340-154-95	Sequence 95, Appl
c	23	39.6	4.1	2127	5	PCT-US95-09338-95	Sequence 95, Appl
c	24	39.6	4.1	2127	5	PCT-US95-09339-95	Sequence 95, Appl
	25	39.4	4.0	2960	3	US-08-913-842-3	Sequence 3, Appli
	26	39.4	4.0	8700	2	US-08-392-625-16	Sequence 16, Appl
	27	39.4	4.0	8700	2	US-08-466-961A-16	Sequence 16, Appl
	28	39.4	4.0	8700	2	US-08-645-193B-18	Sequence 18, Appl
c	29	39.2	4.0	658	4	US-08-998-416-595	Sequence 595, App
	30	39	4.0	688	4	US-08-998-416-972	Sequence 972, App
c	31	38.6	4.0	246240	2	US-08-724-394A-20	Sequence 20, Appl
c	32	38.6	4.0	246240	2	US-08-724-394A-21	Sequence 21, Appl
c	33	38.6	4.0	246240	2	US-08-724-394A-22	Sequence 22, Appl
c	34	38.4	3.9	665	2	US-08-883-795A-36	Sequence 36, Appl
	35	38.4	3.9	2435	4	US-09-306-593-1	Sequence 1, Appli
	36	38	3.9	2950	5	PCT-US93-08386-7	Sequence 7, Appli
	37	38	3.9	2968	5	PCT-US93-08386-1	Sequence 1, Appli
	38	37.8	3.9	899	1	US-08-353-341-4	Sequence 4, Appli
	39	37.6	3.9	658	4	US-08-998-416-193	Sequence 193, App
	40	37.6	3.9	676	4	US-08-998-416-1014	Sequence 1014, Ap
	41	37.6	3.9	677	4	US-08-998-416-920	Sequence 920, App
c	42	37.6	3.9	1511	1	US-07-991-867B-8	Sequence 8, Appli
c	43	37.6	3.9	1511	1	US-08-107-755A-8	Sequence 8, Appli
c	44	37.6	3.9	1511	2	US-08-544-332-8	Sequence 8, Appli
c	45	37.4	3.8	636	4	US-08-998-416-1137	Sequence 1137, Ap

ALIGNMENTS

```
RESULT 1
US-08-988-251-1/c
; Sequence 1, Application US/08988251
; Patent No. 6013497
;  GENERAL INFORMATION:
;    APPLICANT:  Wallis, Nicola G.
;    TITLE OF INVENTION:  MurE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  4000 Bell Atlantic Tower, 1717 Arch Stre
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  US
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/988,251
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dickinson, Todd Q
;      REGISTRATION NUMBER:  28,354
;      REFERENCE/DOCKET NUMBER:  GM10094
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-994-2252
;      TELEFAX:  215-994-2222
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2600 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-988-251-1


  Query Match             44.8%;  Score 437;  DB 3;  Length 2600;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-97;
  Matches  437;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atggttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatg 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    557 ATGGTTACTCAAAATAAAAAGATATTGATTATTACTGGCTCATTCGGTAACGGTCATATG 498

Qy     61 caagttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 CAAGTTACACAGAGTATCGTTAATCAACTTAATGATATGAATCTAGACCATTTAAGCGTC 438

Qy    121 attgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatgg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 ATTGAGCACGATTTATTTATGGAAGCTCATCCAATTTTGACTTCTATTTGTAAAAAATGG 378

Qy    181 tatatcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgccca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 TATATCAATAGCTTTAAATATTTTAGAAATATGTACAAAGGGTTTTATTACAGCCGCCCA 318

Qy    241 gataaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 GATAAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTTATTG 258

Qy    301 ataaaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtacta 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 ATAAAAGAAAAGCCAGATTTAATATTATTAACGTTTCCTACACCAGTTATGTCGGTACTA 198

Qy    361 actgagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 ACTGAGCAATTTAACATTAATATTCCAGTTGCTACAGTGATGACAGACTATCGCTTACAT 138

Qy    421 aaaaactggattacgcc 437
          |||||||||||||||||
Db    137 AAAAACTGGATTACGCC 121


RESULT   2
US-09-386-048-1/c
; Sequence 1, Application US/09386048
; Patent No. 6103507
; GENERAL INFORMATION:
;  APPLICANT: Wallis, Nicola
;  APPLICANT: Shilling, Lisa
;  APPLICANT: Zabeierowski, Susan
;  TITLE OF INVENTION: MurE
;  FILE REFERENCE: GM10094-D1
;  CURRENT APPLICATION NUMBER: US/09/386,048
;  CURRENT FILING DATE: 1999-08-30
;  EARLIER APPLICATION NUMBER: US 08/988,251
;  EARLIER FILING DATE: 1997-12-10
;  EARLIER APPLICATION NUMBER: US 60/059,956
;  EARLIER FILING DATE: 1997-09-25
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2600
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-09-386-048-1


  Query Match             44.8%;  Score 437;  DB 4;  Length 2600;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-97;
  Matches  437;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atggttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatg 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 ATGGTTACTCAAAATAAAAAGATATTGATTATTACTGGCTCATTCGGTAACGGTCATATG 498

Qy     61 caagttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 CAAGTTACACAGAGTATCGTTAATCAACTTAATGATATGAATCTAGACCATTTAAGCGTC 438

Qy    121 attgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatgg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 ATTGAGCACGATTTATTTATGGAAGCTCATCCAATTTTGACTTCTATTTGTAAAAAATGG 378

Qy    181 tatatcaatagctttaaatattttagaaatatgtacaaagggttttattacagccgccca 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 TATATCAATAGCTTTAAATATTTTAGAAATATGTACAAAGGGTTTTATTACAGCCGCCCA 318

Qy    241 gataaactagacaaatgtttttacaaatactatggacttaataagttaattaatttattg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 GATAAACTAGACAAATGTTTTTACAAATACTATGGACTTAATAAGTTAATTAATTTATTG 258

Qy    301 ataaaagaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtacta 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    257 ATAAAAGAAAAGCCAGATTTAATATTATTAACGTTTCCTACACCAGTTATGTCGGTACTA 198

Qy    361 actgagcaatttaacattaatattccagttgctacagtgatgacagactatcgcttacat 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 ACTGAGCAATTTAACATTAATATTCCAGTTGCTACAGTGATGACAGACTATCGCTTACAT 138

Qy    421 aaaaactggattacgcc 437
          |||||||||||||||||
Db    137 AAAAACTGGATTACGCC 121


RESULT   3
US-08-446-855A-1
; Sequence 1, Application US/08446855A
; Patent No. 5849573
;  GENERAL INFORMATION:
;    APPLICANT:  Stewart, Thomas S
;    APPLICANT:  Flores, Maria V
;    APPLICANT:  O'Sullivan, William J
;    TITLE OF INVENTION:  Nucleotide sequence encoding carbamoyl
;    TITLE OF INVENTION:  phosphate synthetase II
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye PC
;      STREET:  1100 No. 5849573th Glebe Road, 8th Floor
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  USA
;      ZIP:  22201-4714
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,855A
;      FILING DATE:  06-Jul-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mitchard, Leonard C
;      REGISTRATION NUMBER:  29.009
;      REFERENCE/DOCKET NUMBER:  47-80
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-816-4000
;      TELEFAX:  703-816-4100
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8920 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic
US-08-446-855A-1


  Query Match             4.8%;  Score 46.4;  DB 2;  Length 8920;
  Best Local Similarity   48.7%;  Pred. No. 0.015;
  Matches  155;  Conservative    0;  Mismatches  161;  Indels    2;  Gaps    1;

Qy     69 acagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattgagca 128
          | | |  ||   ||||  |  ||   ||||  || || |   ||      | | | |
Db   8502 ATATATAATATATAATATATATATATATATATATATATATTTATTTATTTAATTATAT 8561

Qy    129 cgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtatatcaa 188
```

```
Db   8562 TTACGTTTAAATATTAATAAATGTTTTTATTAAATATGATCATTAATTTATATTGATTTA 8621

Qy    189 tagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagataaact 248
Db   8622 TTTTTTTATAAATTTTTGTTATATATACAAATTTTATTTATTCACTCATATGTATAAACC 8681

Qy    249 agacaaatgtttttacaaatactatggacttaataagttaattaatttattgata--aaa 306
Db   8682 AAAATGGTTTTTTCAATTTACAAATAATTTTATAATTTTAATAAATTTATTAATTATAAA 8741

Qy    307 gaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactgag 366
Db   8742 AAAAATAAAAATATATAAACATTAAAATGTATAAATTCTTTTAATTATATAATAATTTAT 8801

Qy    367 caatttaacattaatatt 384
Db   8802 AAATGTTATGATTTTTTT 8819


RESULT   4
US-09-150-741-1
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
;  APPLICANT: Stewart et al.
;  TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
;  TITLE OF INVENTION: Synthetase II
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/150,741
;  CURRENT FILING DATE: 1998-09-10
;  EARLIER APPLICATION NUMBER: PL6380
;  EARLIER FILING DATE: 1992-12-16
;  EARLIER APPLICATION NUMBER: AU93/00617
;  EARLIER FILING DATE: 1993-12-02
;  EARLIER APPLICATION NUMBER: 08/446,855
;  EARLIER FILING DATE: 1995-07-06
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 8920
;   TYPE: DNA
;   ORGANISM: Plasmodium falciparum
US-09-150-741-1


  Query Match             4.8%;  Score 46.4;  DB 4;  Length 8920;
  Best Local Similarity   48.7%;  Pred. No. 0.015;
  Matches  155;  Conservative    0;  Mismatches  161;  Indels    2;  Gaps    1;

Qy     69 acagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattgagca 128
Db   8502 atatataatatatatatatatatatatatatatatatatatatttatttatttaattatat 8561

Qy    129 cgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtatatcaa 188
Db   8562 ttacgtttaaatattaataaatgtttttattaaatatgatcattaatttatattgattta 8621

Qy    189 tagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagataaact 248
Db   8622 ttttttataaatttttgttatatatacaaatttttatttattcactcatatgtataaacc 8681

Qy    249 agacaaatgtttttacaaatactatggacttaataagttaattaatttattgata--aaa 306
Db   8682 aaaatggttttttcaatttacaaataatttttataatttttaataaatttattaattataaa 8741

Qy    307 gaaaagccagatttaatattattaacgtttcctacaccagttatgtcggtactaactgag 366
Db   8742 aaaaataaaaatatatataaacattaaaatgtataaattcttttaattatataataatttat 8801

Qy    367 caatttaacattaatatt 384
Db   8802 aaatgttatgatttttttt 8819


RESULT   5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;   APPLICANT:  DORNER, F.
;   APPLICANT:  SCHEIFLINGER, F.
;   APPLICANT:  FALKNER, F. G.
;   TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;   NUMBER OF SEQUENCES:  52
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Foley & Lardner
;     STREET:  1800 Diagonal Road, Suite 500
;     CITY:  Alexandria
;     STATE:  VA
;     COUNTRY:  USA
;     ZIP:  22313-0299
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/232,463
;     FILING DATE:
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US/07/935,313
;     FILING DATE:
;     APPLICATION NUMBER:  EP 91 114 300.6
;     FILING DATE:  26-AUG-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  BENT, Stephen A.
;     REGISTRATION NUMBER:  29,768
;     REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (703)836-9300
;     TELEFAX:  (703)683-4109
;     TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  7218 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   IMMEDIATE SOURCE:
;     CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             4.6%;  Score 44.8;  DB 1;  Length 7218;
  Best Local Similarity   6.3%;  Pred. No. 0.034;
  Matches   25;  Conservative  203;  Mismatches  170;  Indels    0;  Gaps    0;

Qy    372 taacattaatattccagttgctacagtgatgacagactatcgcttacataaaaactggat 431
Db   1479 TTACCTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRR 1420

Qy    432 tacgccgtattcaacaagatattatgtggcaacaaaagaaacgaaacaagacttcataga 491
Db   1419 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1360

Qy    492 cgtaggtattgatccttcaacagttaaagtgacaggtattcctattgataacaaatttga 551
Db   1359 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1300

Qy    552 aacgcctattaatcaaaagcagtggttaatagacaacaacttagatccagataagcaaac 611
Db   1299 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1240
```

```
Qy    612 tattttaatgtcagctggtgcatttggtgtatctaaaggttttgacacgatgattactga 671
          :   :: :  ::  :: : :   :: : :   :::::    :: : :: ::  :  ::
Db   1239 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1180

Qy    672 tatattagcgaaaagtgcaaatgcacaagtagtttatgatttgtggtaagagcaaagagct 731
          : :  :: :::::: : ::: : : ::: ::  : ::   : :: ::::: ::::::
Db   1179 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1120

Qy    732 aaagcgttctttaacagctaagtttaaattaacgagaa 769
          :::: :      :: ::  :::   :::  :: :::::
Db   1119 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1082


RESULT   6
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13


  Query Match             4.5%;  Score 43.8;  DB 2;  Length 19124;
  Best Local Similarity   49.1%;  Pred. No. 0.075;
  Matches  143;  Conservative    0;  Mismatches  147;  Indels    1;  Gaps    1;

Qy      6 tactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatgcaagt 65
          ||   | |||||||  | | |  | | || |   ||||    ||    |  ||   | ||
Db   5346 TAAAAATAATAAAATAAGAATACTGAATAATAAGTCATATTATACATTTTTTAAAAATGT 5405

Qy     66 tacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtcattga 125
           ||| |       |   | |  |  || |||  |||    |  ||||||   | |  ||
Db   5406 AACATAATTACAAATACGTAACATGTATTATAGAAATAATAAGAATTTAA-TATTAAGGA 5464
```

```
Qy    126 gcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtatat 185
           |  ||  || | |  ||    ||    || || |   |  |  || |  ||  |||||
Db   5465 TAAATATAAATATTTAAAATTATATTTTTTTATGTCAATTTATGTTATATTATATTATAT 5524

Qy    186 caatagctttaaatattttagaaatatgtacaaagggttttattacagccgcccagataa 245
           || |   |||  | ||| | |||||| || | |   | | || | |       || |||
Db   5525 TAACATGATTAGTTTTTTGAAAAATATTTAAATATCATATAATAATAATAAATTAGTTAA 5584

Qy    246 actagacaaatgtttttacaaatactatggacttaataagttaattaattt 296
           | ||     || |  |   ||||||||     | | ||  | |  |||| |
Db   5585 AATAATAGTATTTCATACAAAATACTAACTTATAAGTATATCATATAATAT 5635


RESULT   7
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13


  Query Match             4.3%;  Score 42;  DB 2;  Length 19124;
  Best Local Similarity   47.1%;  Pred. No. 0.2;
  Matches  129;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy    132 tttatttatggaagctcatccaattttgacttctatttgtaaaaaatggtatatcaatag 191
          | ||| |||  |     ||   | |     | |||||   | |  |   |||||  |||
Db   7095 TATATATATATATATATATATTAATAGTTGTACTATTATAATATTACAATATATGTATAT 7036

Qy    192 ctttaaatattttagaaatatgtacaaagggttttattacagccgcccagataaactaga 251
           || ||| | |       || |  | |   |  || |         | |||   |  |
```

```
Db   7035 GTTAAAAAAATAATATTTAATATGTATATAATAATAATTATTAGTTTTATATATTTTTAA 6976

Qy    252 caaatgtttttacaaatactatggacttaataagttaattaatttattgataaaagaaaa 311
          |||  | | || | ||| ||   | || ||||   | |||| |  |  |  ||| ||||
Db   6975 AAAAAATATATATATATATTAATAAATTTATAATAAATTTAAATATTCTAACAAAAAAAA 6916

Qy    312 gccagatttaatattattaacgtttcctacaccagttatgtcggtactaactgagcaatt 371
            | | |  | |||||  |||       | |||| |   || | | | |   ||
Db   6915 AATATAATCAGAAATATTATATTTTATGTATTCCTTTATTTATCTATTTAATTATATATA 6856

Qy    372 taacattaatattccagttgctacagtgatgaca 405
          ||  |||  | ||   ||  |  | | |  | |
Db   6855 TATTATTTTTTTTATGTTTTATTTATTAAGTAAA 6822


RESULT   8
US-08-749-391-1
; Sequence 1, Application US/08749391
; Patent No. 5948667
; GENERAL INFORMATION:
;    APPLICANT:  Cheng, Kuo-Joan
;    APPLICANT:  Selinger, Leonard B.
;    APPLICANT:  Liu, Jin-Hao
;    APPLICANT:  Hu, Youji
;    APPLICANT:  Forsberg, Cecil W.
;    APPLICANT:  Moloney, Maurice M.
;    TITLE OF INVENTION:  A Xylanase Obtained From an
;    TITLE OF INVENTION:  Anaerobic Fungus
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80803
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/749,391
;      FILING DATE:  13-NOV-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Donna M. Ferber
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  93-96
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2058 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Neocallimastix patriciarum
;      STRAIN:  27
;    IMMEDIATE SOURCE:
;      LIBRARY:  genomic DNA library
;      CLONE:  pNspX-06
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  301..1755
US-08-749-391-1
```

```
  Query Match             4.2%;  Score 41;  DB 2;  Length 2058;
  Best Local Similarity   46.7%;  Pred. No. 0.21;
  Matches  163;  Conservative    0;  Mismatches  185;  Indels    1;  Gaps    1;

Qy      1 atggttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatg 60
          ||  ||  |||||| ||  |   ||  |  |  | |   |  ||   |   |  | |
Db      9 ATAATTGTTCAAAAAAAGTAATAATAAAAAAAAAATTTTTTTTTTTTTTTTGGGAAAAT 68

Qy     61 caagttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtc 120
            |||   |    | | |   |   ||   | | |   | ||  |  | |||   |
Db     69 TGAGTATAAATAGTTTTTGGTTTACCTTTTTTGGTTTTTCCTTTATTCTTTATAAAGTTA 128

Qy    121 attgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatgg 180
          ||||   |  || |||||  |||||     |  || ||  |  | |||||  | |
Db    129 ATTGTTTAATAATTATTGGTGGAAATATTTAAAAGTTGTATATATATTTTAATATTTATT 188

Qy    181 tatatcaatagctttaaatattttagaaatatgtacaaagggttttat-tacagccgccc 239
            || | |  |||| | |  |  | | | || |  | || || | || ||
Db    189 GGAATTATTTACTTTCACTGGTGGAAACAAATATTAATAGTGTATAATATATTATTAGAA 248

Qy    240 agataaactagacaaatgtttttacaaatactatggacttaataagttaattaatttatt 299
          || | |||  | | || | ||   |  |||       | | ||  ||||||  |    |||
Db    249 AAAGAAAAAAAAAAATTATTACAATTAATTACTATAAATAAAATAGTTAAAAATGAAATT 308

Qy    300 gataaaagaaaagccagatttaatattattaacgtttcctacaccagtt 348
            || ||   |  || |  ||| |||  ||||| |   |||| |  | |
Db    309 TTTACAAATTATTCCTGTATTATTATCTTTAACTTCAACTACTCTTGCT 357


RESULT   9
US-09-390-200-1
; Sequence 1, Application US/09390200
; Patent No. 6137032
; GENERAL INFORMATION:
;    APPLICANT:  Cheng, Kuo-Joan
;    APPLICANT:  Selinger, Leonard B.
;    APPLICANT:  Liu, Jin-Hao
;    APPLICANT:  Hu, Youji
;    APPLICANT:  Forsberg, Cecil W.
;    APPLICANT:  Moloney, Maurice M.
;    TITLE OF INVENTION:  A Xylanase Obtained From an
;    TITLE OF INVENTION:  Anaerobic Fungus
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee, Winner and Sullivan, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80803
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/390,200
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/749,391
;      FILING DATE:  13-NOV-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Donna M. Ferber
;      REGISTRATION NUMBER:  33,878
;      REFERENCE/DOCKET NUMBER:  93-96
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2058 base pairs
;      TYPE:  nucleic acid
```

```
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Neocallimastix patriciarum
;      STRAIN:  27
;    IMMEDIATE SOURCE:
;      LIBRARY:  genomic DNA library
;      CLONE:  pNspX-06
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  301..1755
US-09-390-200-1


  Query Match             4.2%;  Score 41;  DB 4;  Length 2058;
  Best Local Similarity   46.7%;  Pred. No. 0.21;
  Matches  163;  Conservative    0;  Mismatches  185;  Indels    1;  Gaps    1;

Qy      1 atggttactcaaaataaaaagatattgattattactggctcattcggtaacggtcatatg 60
          ||  ||  |||||| ||  |   ||  |  |  | |   |  ||   |    |  | |
Db      9 ATAATTGTTCAAAAAAAGTAATAATAAAAAAAAAATTTTTTTTTTTTTTTTGGGAAAAT 68

Qy     61 caagttacacagagtatcgttaatcaacttaatgatatgaatctagaccatttaagcgtc 120
            |||   |     | | | |   |  ||   | | |   | ||  |  | |||   |
Db     69 TGAGTATAAATAGTTTTTGGTTTACCTTTTTTGGTTTTTCCTTTATTCTTTATAAAGTTA 128

Qy    121 attgagcacgatttatttatggaagctcatccaattttgacttctatttgtaaaaaatgg 180
          ||||   |  | |||||  |||||     |  || ||  |  | |||||  | |
Db    129 ATTGTTTAATAATTATTGGTGGAAATATTTAAAAGTTGTATATATATTTTAATATTTATT 188

Qy    181 tatatcaatagctttaaatattttagaaatatgtacaaagggttttat-tacagccgccc 239
            || | |  |||| | |  |  | | ] || |  | || || | || ||
Db    189 GGAATTATTTACTTTCACTGGTGGAAACAAATATTAATAGTGTATAATATATTATTAGAA 248

Qy    240 agataaactagacaaatgtttttacaaatactatggacttaataagttaattaatttatt 299
          | | |||  | | || | ||   |  |||       | | ||  ||||||  |    |||
Db    249 AAAGAAAAAAAAAAATTATTACAATTAATTACTATAAATAAAATAGTTAAAAATGAAATT 308

Qy    300 gataaaagaaaagccagatttaatattattaacgtttcctacaccagtt 348
            || ||   |  || |  ||| |||  ||||| |   |||| |  | |
Db    309 TTTACAAATTATTCCTGTATTATTATCTTTAACTTCAACTACTCTTGCT 357


RESULT  10
US-08-920-812-6/c
; Sequence 6, Application US/08920812
; Patent No. 5763188
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,812
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Staphylococcus epidermidis
;      STRAIN:  Clinical Isolate SE-22
US-08-920-812-6


  Query Match             4.1%;  Score 40;  DB 1;  Length 8654;
  Best Local Similarity   52.1%;  Pred. No. 0.51;
  Matches   88;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy    195 taaatattttagaaatatgtacaaagggttttattacagccgcccagataaactagacaa 254
          |||| |  |     || |    ||| |  |  |  || |   | ||||   |    |  |
Db   1648 TAAAAAAATCGATGATTTACTTAAACGAGTAAACAACCGTATCACAGAAGCAAATAATGA 1589

Qy    255 atgtttttacaaatactatggacttaataagttaattaatttattgataaaagaaaagcc 314
          |       || ||   ||| ||| | ||  || || ||  || | | || ||||| ||| |
Db   1588 AATGAAGTATAATGGCTAGGGAGTNAAACAGATATTTCTTTCACTTATGAAAGATAAGAC 1529

Qy    315 agatttaatattattaacgtttcctacaccagttatgtcggtactaact 363
            | |||| | || |      ||  |||   ||| || |  || | ||||
Db   1528 TGTTTTACTCTTTTATTAAGTTAATACGATAGTGATCTGAGTCCAAACT 1480


RESULT  11
US-08-920-827-6/c
; Sequence 6, Application US/08920827
; Patent No. 5770375
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,827
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
```

```
;       REGISTRATION NUMBER:  33,547
;       REFERENCE/DOCKET NUMBER:  19036/32420
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  312/474-6300
;       TELEFAX:  312/474-0448
;       TELEX:  25-3856
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  8654 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Staphylococcus epidermidis
;       STRAIN:  Clinical Isolate SE-22
US-08-920-827-6


  Query Match             4.1%;  Score 40;  DB 1;  Length 8654;
  Best Local Similarity   52.1%;  Pred. No. 0.51;
  Matches   88;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy     195 taaatattttagaaatatgtacaaagggttttattacagccgcccagataaactagacaa 254
           |||| |  |     || |    ||| |  |  |  || |   | ||||   |    |  |
Db    1648 TAAAAAAATCGATGATTTACTTAAACGAGTAAACAACCGTATCACAGAAGCAAATAATGA 1589

Qy     255 atgtttttacaaatactatggacttaataagttaattaatttattgataaaagaaaagcc 314
           |      || ||   ||| ||| | ||  || || ||  || | | || ||||| ||| |
Db    1588 AATGAAGTATAATGGCTAGGGAGTNAAACAGATATTTCTTTCACTTATGAAAGATAAGAC 1529

Qy     315 agatttaatattattaacgtttcctacaccagttatgtcggtactaact 363
             | |||| | || |      ||  |||   ||| || |  || | ||||
Db    1528 TGTTTTACTCTTTTATTAAGTTAATACGATAGTGATCTGAGTCCAAACT 1480


RESULT  12
US-08-921-177-6/c
; Sequence 6, Application US/08921177
; Patent No. 5798211
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/921,177
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Staphylococcus epidermidis
;      STRAIN:  Clinical Isolate SE-22
US-08-921-177-6


  Query Match             4.1%;  Score 40;  DB 1;  Length 8654;
  Best Local Similarity   52.1%;  Pred. No. 0.51;
  Matches   88;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy     195 taaatattttagaaatatgtacaaagggttttattacagccgcccagataaactagacaa 254
           |||| |  |     || |    ||| |  |  |  || |   | ||||   |    |  |
Db    1648 TAAAAAAATCGATGATTTACTTAAACGAGTAAACAACCGTATCACAGAAGCAAATAATGA 1589

Qy     255 atgtttttacaaatactatggacttaataagttaattaatttattgataaaagaaaagcc 314
           |      || ||   ||| ||| | ||  || || ||  || | | || ||||| ||| |
Db    1588 AATGAAGTATAATGGCTAGGGAGTNAAACAGATATTTCTTTCACTTATGAAAGATAAGAC 1529

Qy     315 agatttaatattattaacgtttcctacaccagttatgtcggtactaact 363
             | |||| | || |      ||  |||   ||| || |  || | ||||
Db    1528 TGTTTTACTCTTTTATTAAGTTAATACGATAGTGATCTGAGTCCAAACT 1480


RESULT  13
US-08-362-577C-6/c
; Sequence 6, Application US/08362577C
; Patent No. 5807673
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/362,577C
;      FILING DATE:  27-MAR-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
```

```
;    ORIGINAL SOURCE:
;      ORGANISM:  Staphylococcus epidermidis
;      STRAIN:  Clinical Isolate SE-22
US-08-362-577C-6


  Query Match             4.1%;  Score 40;  DB 1;  Length 8654;
  Best Local Similarity   52.1%;  Pred. No. 0.51;
  Matches   88;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy    195 taaatattttagaaatatgtacaaagggttttattacagccgcccagataaactagacaa 254
          |||| |  |     || |    ||| |  |  |  || |   | ||||   |     |  |
Db   1648 TAAAAAAATCGATGATTTACTTAAACGAGTAAACAACCGTATCACAGAAGCAAATAATGA 1589

Qy    255 atgtttttacaaatactatggacttaataagttaattaatttattgataaaagaaaagcc 314
          |      || ||   ||| ||| | ||  || || ||  || | | || ||||| ||| |
Db   1588 AATGAAGTATAATGGCTAGGGAGTNAAACAGATATTTCTTTCACTTATGAAAGATAAGAC 1529

Qy    315 agatttaatattattaacgtttcctacaccagttatgtcggtactaact 363
          | |||| | || |      ||  |||   ||| || |  || | ||||
Db   1528 TGTTTTACTCTTTTATTAAGTTAATACGATAGTGATCTGAGTCCAAACT 1480


RESULT  14
US-08-920-828-6/c
; Sequence 6, Application US/08920828
; Patent No. 5853998
;  GENERAL INFORMATION:
;    APPLICANT:  Ohno, Tsuneya
;    APPLICANT:  Matsuhisa, Akio
;    APPLICANT:  Uehara, Hirotsugu
;    APPLICANT:  Eda, Soji
;    TITLE OF INVENTION:  Probe for Diagnosing Infectious Disease
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/920,828
;      FILING DATE:  29-AUG-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/362,577
;      FILING DATE:  27-MAR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Rin-Laures, Li-Hsien
;      REGISTRATION NUMBER:  33,547
;      REFERENCE/DOCKET NUMBER:  19036/32420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8654 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Staphylococcus epidermidis
;      STRAIN:  Clinical Isolate SE-22
US-08-920-828-6


  Query Match             4.1%;  Score 40;  DB 2;  Length 8654;
  Best Local Similarity   52.1%;  Pred. No. 0.51;
  Matches   88;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;

Qy    195 taaatattttagaaatatgtacaaagggttttattacagccgcccagataaactagacaa 254
          |||| |  |     || |    ||| |  |  |  || |   | ||||   |     |  |
Db   1648 TAAAAAAATCGATGATTTACTTAAACGAGTAAACAACCGTATCACAGAAGCAAATAATGA 1589

Qy    255 atgtttttacaaatactatggacttaataagttaattaatttattgataaaagaaaagcc 314
          |      || ||   ||| ||| | ||  || || ||  || | | || ||||| ||| |
Db   1588 AATGAAGTATAATGGCTAGGGAGTNAAACAGATATTTCTTTCACTTATGAAAGATAAGAC 1529

Qy    315 agatttaatattattaacgtttcctacaccagttatgtcggtactaact 363
          | |||| | || |      ||  |||   ||| || |  || | ||||
Db   1528 TGTTTTACTCTTTTATTAAGTTAATACGATAGTGATCTGAGTCCAAACT 1480


RESULT  15
US-08-505-486-96/c
; Sequence 96, Application US/08505486
; Patent No. 5955573
;  GENERAL INFORMATION:
;    APPLICANT:  Jesse M. Jaynes
;    TITLE OF INVENTION:  UBIQUTIN-LYTIC PEPTIDE FUSION GENE
;    TITLE OF INVENTION:  CONSTRUCTS, PROTEIN PRODUCTS DERIVING THEREFROM, AND
;    TITLE OF INVENTION:  METHODS OF MAKING AND USING SAME
;    NUMBER OF SEQUENCES:  98
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ROTHWELL, FIGG, ERNST & KURZ
;      STREET:  555 Thirteenth Street N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.4 MB STORAGE
;      COMPUTER:  IBM COMPATIBLE
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  WordPerfect 5.1+
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/505,486
;      FILING DATE:  21-JUL-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. 08/279,472
;      FILING DATE:  22-JUL-1994
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  WALKER, BARBARA W.
;      REGISTRATION NUMBER:  35,400
;      REFERENCE/DOCKET NUMBER:  2093-117A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)783-6040
;      TELEFAX:  (202)783-6031
;  INFORMATION FOR SEQ ID NO:  96:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2022
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  DOUBLE STRANDED
;      TOPOLOGY:  LINEAR
;    MOLECULE TYPE:
;      DESCRIPTION:  GENOMIC DNA
US-08-505-486-96


  Query Match             4.1%;  Score 39.6;  DB 2;  Length 2022;
  Best Local Similarity   45.3%;  Pred. No. 0.45;
  Matches  167;  Conservative    0;  Mismatches  201;  Indels    1;  Gaps    1;

Qy     23 tattgattattactggctcattcggtaacggtcatatgcaagttacacagagtatcgtta 82
          |||  | ||| | |   | |   ||||    | | |   || | ||     |  |  | |
Db    386 TATCAAATATAATTTAGTGACGTGGTATAATTAAAAAAAAATTGACCTCTTGCTTTATAA 327
```

```
Qy      83 atcaacttaatgatatgaatctagaccatttaagcgtcattgagcacgatttatttatgg 142
Db     326 GAAATATTATTATTATTACTATATTGCATATAAAACTCAATCAATGTGAATAATAAATGG 267

Qy     143 aagctcatccaattttgacttctatttgtaaaaaatggtatatcaatagctttaaatatt 202
Db     266 GTTTNNNTTGTGTCTGTCATTATCATTATGAAAAACATTGAAGTTGCTTACTATTCAATT 207

Qy     203 ttagaaatatgtacaaagggttttattacagccgcccagataaactagacaaatgt-ttt 261
Db     206 TGGGTATATGTAATTGATGACCAANATCTTGTCATCCTCATCAACAAAACTAATGTAAAT 147

Qy     262 tacaaatactatggacttaataagttaattaatttattgataaaagaaaagccagattta 321
Db     146 NTTTAATAAAACTTTTTGAAGNTGNAAATTGTCTTGGATATAAAGTGAATTTAAAAAATC 87

Qy     322 atattattaacgtttcctacaccagttatgtcggtactaactgagcaatttaacattaat 381
Db      86 ATAAAATTGAGATTTCAAATATTAGTTTAATGTAGAATTATTGAGACAATTTATATTTAT 27

Qy     382 attccagtt 390
Db      26 TTATATGTT 18
```

Search completed: June 29, 2001, 11:47:46
Job time: 6678 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 29, 2001, 08:59:02 ; Search time 35.83 Seconds
(without alignments)
661.568 Million cell updates/sec

Title: US-09-668-788-4
Perfect score: 2048
Sequence: 1 MVTQNKKILIITGSFGNGHM.....SSQPQEIYGKVPLYARFFVK 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	591.5	28.9	382	20	AAY33444	B. subtilis glycos
2	296.5	14.5	468	21	AAB19042	Amino acid sequenc
3	284	13.9	533	21	AAB19041	Amino acid sequenc
4	283	13.8	422	19	AAW41148	Monogalactosyldiac
5	283	13.8	525	21	AAB19040	Amino acid sequenc
6	279.5	13.6	522	21	AAB19039	Amino acid sequenc
7	272	13.3	492	21	AAG42414	Arabidopsis thalia
8	272	13.3	551	21	AAG42413	Arabidopsis thalia
9	252	12.3	404	21	AAG42415	Arabidopsis thalia
10	139	6.8	374	20	AAW88310	E. coli colitose o
11	119	5.8	352	19	AAW51348	Streptococcus pneu
12	109.5	5.3	534	13	AAR26154	HUG-Br2. Homo sap
13	106	5.2	530	21	AAY78935	Human UDP-glucuron
14	102.5	5.0	934	16	AAR76063	Human MSH2 protein
15	102.5	5.0	934	16	AAR75411	Human MSH2. Homo
16	102.5	5.0	934	18	AAW09034	Mismatch repair pr
17	102	5.0	530	19	AAW47126	Uridine diphospho-
18	101.5	5.0	367	21	AAY81719	Streptococcus pneu
19	98.5	4.8	2366	17	AAR95011	C. difficile toxin
20	98.5	4.8	2366	19	AAW68388	Clostridium diffic
21	98	4.8	510	14	AAR68893	Japanese Black Pin
22	96.5	4.7	533	13	AAR26153	HUG-Br1. Homo sap
23	96	4.7	662	18	AAW20999	H. pylori cytoplas
24	95.5	4.7	1435	20	AAY49070	PolC gene product
25	95.5	4.7	1435	22	AAB31934	Amino acid sequenc
26	95	4.6	467	21	AAG43513	Arabidopsis thalia
27	94.5	4.6	484	21	AAG11789	Arabidopsis thalia
28	94.5	4.6	484	21	AAG21988	Arabidopsis thalia
29	92.5	4.5	427	21	AAG43514	Arabidopsis thalia
30	92.5	4.5	1151	20	AAW95039	Human N-arginine d
31	92	4.5	648	19	AAW75910	Helicobacter methi
32	92	4.5	1829	18	AAW29322	DNA polymerase wit
33	90.5	4.4	503	21	AAY81750	Streptococcus pneu
34	90	4.4	1802	21	AAB18217	Plasmodium falcipa
35	89	4.3	350	19	AAW62736	Streptococcus pneu
36	89	4.3	1049	21	AAB42125	Human ORFX ORF1889
37	89	4.3	1092	19	AAW41602	Staphylococcus epi
38	89	4.3	1174	21	AAB19190	Lipid metabolism p
39	88.5	4.3	366	21	AAG43515	Arabidopsis thalia
40	88	4.3	487	19	AAW98305	H. pylori GHPO 479
41	88	4.3	666	18	AAW13729	Herbicide-resistan
42	87.5	4.3	487	22	AAY72635	Exophiala spinifer
43	87.5	4.3	873	21	AAG48384	Arabidopsis thalia
44	87.5	4.3	2120	21	AAY81710	Streptococcus pneu
45	87	4.2	397	21	AAG43120	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAY33444
ID AAY33444 standard; Protein; 382 AA.
XX
AC AAY33444;
XX
DT 13-DEC-1999 (first entry)
XX
DE B. subtilis glycosyl transferase catalytic domain.
XX
KW Catalytic domain; glycosyl transferase; processive activity; detergent;
KW oligosaccharide glycolipid; 3-oligoglucosyl-1,2-diacylglycerol;
KW food emulsifier; polymer modifier; glycosyl transferase.
XX
OS Bacillus subtilis.
XX
PN DE19819958-A1.
XX
PD 30-SEP-1999.
XX
PF 05-MAY-1998; 98DE-1019958.
XX
PR 25-MAR-1998; 98DE-1013017.
XX
PA (GVSE-) GVS GES ERWERB & VERW LANDWIRTSCHAFTLICH.
PA (BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.
XX
PI Wolter FP, Jorasch P, Heinz E, Zaehringer U;
XX
DR WPI; 1999-552364/47.
DR N-PSDB; AAZ23386.
XX
PT New glycosyl transferase protein, useful for producing glycolipids -
XX

```
PS   Example 1; Page 7; 8pp; German.
XX
CC   This invention describes a novel protein (I) comprising identical or
CC   different catalytic domains of glycosyl transferases and which has
CC   processive activity (i.e. builds up oligosaccharide glycolipids by
CC   successive addition of glucose units). (I) can be used to produce
CC   3-oligoglucosyl-1,2-diacylglycerols useful as food emulsifiers, polymer
CC   modifiers or detergents. This sequence represents a Bacillus subtilis
CC   glycosyl transferase catalytic domain which is used in the method of
CC   the invention.
XX
SQ   Sequence   382 AA;


  Query Match             28.9%;  Score 591.5;  DB 20;  Length 382;
  Best Local Similarity   35.7%;  Pred. No. 8.4e-48;
  Matches  136;  Conservative   77;  Mismatches  131;  Indels   37;  Gaps   10;

Qy       5 NKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINS 64
Db       4 nkrvliltanygnghvqvaktlyeqcvrlgfqhvtv--snlyqesnpivsevtqylylks 61

Qy      65 FKYFRNMYKGFYYSRPDKLDKCFYK-----YYGL-NKLINLLIKE-KPDLILLTFPTPVM 117
Db      62 fsigkqfyrlfyyg----vdkiynkrkfniyfkmgnkrlgelvdehqpdiiintfpmivv 117

Qy     118 SVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGI 177
Db     118 peyrrrtgrviptfnvmtdfclhkiwvhenvdkyyvatdyvkekllleigthpsnvkitgi 177

Qy     178 PIDNKFET-----PINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSANA 232
Db     178 pirpqfeesmpvgpiykky-----nlspnkkvllimagahgvlknvkelcen-lvkddqv 231

Qy     233 QVVMICGKSKELKRSLTA-------KFKLTRMYLILGYTKHMNEWMASSQLMITKPGGIT 285
Db     232 qvvvvcgkntalkeslsaleaengdklk------vlgyveridelfritdcmitkpggit 285

Qy     286 ITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTNGNEQLTNM 345
Db     286 lteataigvpvilykpvpgqekenanffedrgaaivvnrheeilesvtslladedtlhrm 345

Qy     346 ISTMEQDKIKYATQTICRDLL 366
Db     346 kknikdlhlansseviledil 366


RESULT   2
AAB19042
ID   AAB19042 standard; protein; 468 AA.
XX
AC   AAB19042;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Amino acid sequence of monogalactosyl-diacylglycerol (MGDG) synthase.
XX
KW   Monogalactosyl-diacylglycerol synthase; MGDG synthase; Toxoplasma;
KW   apicomplex parasite; herbicide; antiparasitic; Plasmodium; Eimeria;
KW   acquired immune deficiency syndrome; coccidiosis.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200056919-A1.
XX
PD   28-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-FR00658.
XX
PR   19-MAR-1999;   99FR-0003434.
XX
PA   (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI   Marechal E,  Block M,  Joyard J,  Douce R;
XX
DR   WPI; 2000-602227/57.
XX
PT   Use of monogalactosyl diacylglycerol synthase for identifying its
PT   specific inhibitors, potentially useful as antiparasitic agents and
PT   herbicides -
XX
PS   Example 2; Fig 2; 33pp; French.
XX
CC   The present sequence represents a monogalactosyl-diacylglycerol (MGDG)
CC   synthase. MGDG is present in all plastids tested and is essential for
CC   cell survival, but is not present in other membrane structures, or in
CC   animal cells, and so represents a specific target. MGDG synthase or a
CC   plastid membrane isolated from plants is used to select and screen for
CC   specific inhibitors of MGDG synthase. These inhibitors are suitable as
CC   active agents against apicomplex parasites or as herbicides. The
CC   inhibitors are used as antiparasitic agents, especially against
CC   Plasmodium, Toxoplasma (particularly in patients with acquired immune
CC   deficiency syndrome) or Eimeria (coccidiosis in cattle or poultry), and
CC   as herbicides.
XX
SQ   Sequence   468 AA;


  Query Match             14.5%;  Score 296.5;  DB 21;  Length 468;
  Best Local Similarity   25.4%;  Pred. No. 1.1e-19;
  Matches  100;  Conservative   78;  Mismatches  173;  Indels   43;  Gaps   14;

Qy       4 QNKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYIN 63
Db      67 rtknvlilmsdtggghrasaeairdafkiefgdkyrvivkdvwke--------ytgwpln 118

Qy      64 ----SFKYF---RNMYK-GFYYSRPDKLDKCFY----KYYGLNKLINLLIKEKPDLILLT 111
Db     119 dmersykfmvkhvqlwkvafhstspkwihscylaaiaayya-keveaglmeykpeiiisv 177

Qy     112 FP----TPVMSVLTEQFNINIPVATVMTDYR-LHKNWITPYSTRYYVATKETKQDFIDVG 166
Db     178 hplmqhiplwvlkwqelqkrvlfvtvitdlntchptwfhpgvnrcycpsqevakralfdg 237

Qy     167 IDPSTVKVTGIPIDNKFETPINQKQWL-IDNNLDPDKQTILMSAGAFGV------SKGFD 219
Db     238 ldesqvrvfglpvrpsfaravlvkddlrkelemdqdlravllmgggegmgpvketakale 297

Qy     220 TMITDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLMI 278
Db     298 eflydkenrkpigqmvvicgrnkklasaleaidwkip--vkvrgfetqmekwmgacdcii 355

Qy     279 TKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIV----AS 334
Db     356 tkagpgtiaeslirslpiilndyipgqekgnvpyvvengagvftrspketarivgewfst 415

Qy     335 LTNGNEQLTNMISTMEQDKIKYATQTICRDLLDL 368
Db     416 ktdeleqtsdnarklaqpe---avfdivkdidel 446


RESULT   3
AAB19041
ID   AAB19041 standard; protein; 533 AA.
XX
AC   AAB19041;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Amino acid sequence of monogalactosyl-diacylglycerol (MGDG) synthase.
XX
KW   Monogalactosyl-diacylglycerol synthase; MGDG synthase; Toxoplasma;
KW   apicomplex parasite; herbicide; antiparasitic; Plasmodium; Eimeria;
KW   acquired immune deficiency syndrome; coccidiosis.
XX
OS   Arabidopsis thaliana.
```

XX
PN WO200056919-A1.
XX
PD 28-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-FR00658.
XX
PR 19-MAR-1999; 99FR-0003434.
XX
PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI Marechal E, Block M, Joyard J, Douce R;
XX
DR WPI; 2000-602227/57.
XX
PT Use of monogalactosyl diacylglycerol synthase for identifying its
PT specific inhibitors, potentially useful as antiparasitic agents and
PT herbicides -
XX
PS Example 2; Fig 2; 33pp; French.
XX
CC The present sequence represents a monogalactosyl-diacylglycerol (MGDG)
CC synthase. MGDG is present in all plastids tested and is essential for
CC cell survival, but is not present in other membrane structures, or in
CC animal cells, and so represents a specific target. MGDG synthase or a
CC plastid membrane isolated from plants is used to select and screen for
CC specific inhibitors of MGDG synthase. These inhibitors are suitable as
CC active agents against apicomplex parasites or as herbicides. The
CC inhibitors are used as antiparasitic agents, especially against
CC Plasmodium, Toxoplasma (particularly in patients with acquired immune
CC deficiency syndrome) or Eimeria (coccidiosis in cattle or poultry), and
CC as herbicides.
XX
SQ Sequence 533 AA;

Query Match 13.9%; Score 284; DB 21; Length 533;
Best Local Similarity 25.5%; Pred. No. 2.1e-18;
Matches 103; Conservative 66; Mismatches 185; Indels 50; Gaps 13;

```
Qy      6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSF 65
          ||:||:    | ||    ::|    |    |   |   ||: :  |         |  |
Db    142 kkvlilmsdtggghrasaeairaafnqefgdeyqvfitdlwtdhtp--------wpfnql 193

Qy     66 KYFRN-------MYKGFYYSRPDKL----DKCFYKYYGLNKLINLLIKEKPDLILLTFP- 113
              |       ::|  ||    ::    :       :   ::   |:| :||:|:   |
Db    194 prsynflvkhgtlwkmtyygtsprivhqsnfaatstfiareiaqglmkyqpdiiisvhpl 253

Qy    114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
              |:  : ::     |   ||:||    |  |     || |  : |  :     |::
Db    254 mqhvplrvlrskgllkkivfttvitdlstchptwfhklvtrcycpstevakraqkaglet 313

Qy    170 STVKVTGIPIDNKFETPINQKQWL-----IDNNLDPDKQTILMSAGAFGV------SKGF 218
          | :|| |:|:   |  |:  |  |     :| ||       :|:  |  |:       :: .
Db    314 sqikvyglpvrpsfvkpvrpkvelrrelgmdenl----pavllmgggegmgpieataral 369

Qy    219 DTMITDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLM 277
          : |     |  ||::|||::|:|:  |::  :|:     : |:   | | | :   :
Db    370 adalydknlgeavgqvliicgrnkklqsklssldwkip--vqvkgfitkmeecmgacdci 427

Qy    278 ITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTN 337
          ||| |  || |   | :|:|      |||  |  |  | | || : :|:|  ||||
Db    428 itkagpgtiaeamirglpiilngyiagqeagnvpyvvengcgkfskspkeiskivadwfg 487

Qy    338 GNEQLTNMISTMEQDKIKYATQ----TICRDLLDLI-GHSSQPQ 376
              :  | |: :: |       |  |: :|:   :| ||
Db    488 ---paskeleimsqnalrlakpeavfkivhdmhelvrkknslpq 528
```

RESULT 4
AAW41148
ID AAW41148 standard; Protein; 422 AA.
XX

AC AAW41148;
XX
DT 05-MAY-1998 (first entry)
XX
DE Monogalactosyldiacylglycerol synthetase sequence.
XX
KW Monogalactosyldiacylglycerol synthetase; MGDG; lipid production;
KW cucumber.
XX
OS Cucumis sativus.
XX
PN JP10014579-A.
XX
PD 20-JAN-1998.
XX
PF 02-JUL-1996; 96JP-0172337.
XX
PR 02-JUL-1996; 96JP-0172337.
XX
PA (KIRI) KIRIN BREWERY KK.
XX
DR WPI; 1998-138241/13.
DR N-PSDB; AAV12734.
XX
PT New mono:galactosyl-di:acyl-glycerol synthase - useful for, e.g.
PT increasing lipid production in transformed host
XX
PS Claim 1; Page 8; 13pp; Japanese.
XX
CC This sequence represents the monogalactosyldiacylglycerol synthase (MGDG)
CC of the invention. This sequence was isolated from cucumber. MGDG may be
CC used to increase lipid production in an organism.
XX
SQ Sequence 422 AA;

Query Match 13.8%; Score 283; DB 19; Length 422;
Best Local Similarity 25.3%; Pred. No. 1.9e-18;
Matches 105; Conservative 65; Mismatches 185; Indels 60; Gaps 13;

```
Qy      6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSF 65
          |::||:    | ||    ::|    |:   ::  |   ||: :  |         |  |
Db     33 krvlilmsdtggghrasaeaikaafneefgnnyqvfitdlwtdhtp--------wpfnql 84

Qy     66 KYFRN-------MYKGFYYSRPDKL----DKCFYKYYGLNKLINLLIKEKPDLILLTFP- 113
              |       ::|  ||    |:    :       :   ::   |:| :||:|:   |
Db     85 prsynflvkhgtlwkmtyyvtapkvihqsnfaatstfiarevakglmkyrpdiiisvhpl 144

Qy    114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
              |:  : ::     |   ||:||    |  |     || |  : |  :     |: |
Db    145 mqhvpirilrskgllnkivfttvvtdlstchptwfhklvtrcycpstevakralkaglqp 204

Qy    170 STVKVTGIPIDNKFETPIN-----QKQWLIDNNLDPDKQTILMSAGAFGV------SKGF 218
          | :|| |:|:   |  ||       :|:  :| ||       :|:  |  |:       :|
Db    205 sklkvfglpvrpsfvkpirpkielrkelgmdenl----pavllmgggegmgpieatakal 260

Qy    219 DTMITDILAKSANAQVVMICGKSKELK---RSLTAKFKLTRMYLILGYTKHMNEWMASSQ 275
              : |        ||::||| :|:|    ||:  |  :     : |:   | | | :
Db    261 skalydenhgepigqvlvicghnkklagrlrsidwkvpvq----vkgfvtkmeecmgacd 316

Qy    276 LMITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASL 335
              :||| |  || |   | :|:|  :   |||  |  |  | | || : :|:|   |||
Db    317 ciitkagpgtiaeamirglpiilndyiagqeagnvpyvvengcgkfskspkeianivakw 376

Qy    336 TNGNEQLTNMISTMEQDKIKYATQ----TICRDLLDLIGHSSQPQEIYGKVPLYA 386
              : :  | |: :: |       |  || :|:   |        || |:
Db    377 fgpk---adellimsqnalrlarpdavfkivhdlhelvkqrs-------fvpqys 421
```

RESULT 5
AAB19040
ID AAB19040 standard; protein; 525 AA.
XX

```
AC   AAB19040;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Amino acid sequence of monogalactosyl-diacylglycerol (MGDG) synthase.
XX
KW   Monogalactosyl-diacylglycerol synthase; MGDG synthase; Toxoplasma;
KW   apicomplex parasite; herbicide; antiparasitic; Plasmodium; Eimeria;
KW   acquired immune deficiency syndrome; coccidiosis.
XX
OS   Cucumis sativus.
XX
PN   WO200056919-A1.
XX
PD   28-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-FR00658.
XX
PR   19-MAR-1999;   99FR-0003434.
XX
PA   (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI   Marechal E,  Block M,  Joyard J,  Douce R;
XX
DR   WPI; 2000-602227/57.
XX
PT   Use of monogalactosyl diacylglycerol synthase for identifying its
PT   specific inhibitors, potentially useful as antiparasitic agents and
PT   herbicides -
XX
PS   Example 2; Fig 2; 33pp; French.
XX
CC   The present sequence represents a monogalactosyl-diacylglycerol (MGDG)
CC   synthase. MGDG is present in all plastids tested and is essential for
CC   cell survival, but is not present in other membrane structures, or in
CC   animal cells, and so represents a specific target. MGDG synthase or a
CC   plastid membrane isolated from plants is used to select and screen for
CC   specific inhibitors of MGDG synthase. These inhibitors are suitable as
CC   active agents against apicomplex parasites or as herbicides. The
CC   inhibitors are used as antiparasitic agents, especially against
CC   Plasmodium, Toxoplasma (particularly in patients with acquired immune
CC   deficiency syndrome) or Eimeria (coccidiosis in cattle or poultry), and
CC   as herbicides.
XX
SQ   Sequence   525 AA;

  Query Match            13.8%;  Score 283;  DB 21;  Length 525;
  Best Local Similarity  25.3%;  Pred. No. 2.6e-18;
  Matches  105;  Conservative   65;  Mismatches  185;  Indels   60;  Gaps   13;

Qy       6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSF 65
           |::||:    | ||    ::|    |:   ::  |   ||: :  |        |  |
Db     136 krvlilmsdtggghrasaeaikaafneefgnnyqvfitdlwtdhtp--------wpfnql 187

Qy      66 KYFRN-------MYKGFYYSRPDKL----DKCFYKYYGLNKLINLLIKEKPDLILLTFP- 113
             |       ::|  ||    |:    :      :   ::   |:| :||:|:   |
Db     188 prsynflvkhgtlwkmtyyvtapkvihqsnfaatstfiarevakglmkyrpdiiisvhpl 247

Qy     114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
               |:  : ::     |   ||:||    |  |     || |  : |  :  :  |: |
Db     248 mqhvpirilrskgllnkivfttvvtdlstchptwfhklvtrcycpstevakraltaglqp 307

Qy     170 STVKVTGIPIDNKFETPIN-----QKQWLIDNNLDPDKQTILMSAGAFGV------SKGF 218
           | :|| |:|:   |  ||      :|:  :| ||      :|:  |  |:      :|
Db     308 sklkvfglpvrpsfvkpirpkielrkelgmdenl----pavllmgggegmgpieatakal 363

Qy     219 DTMITDILAKSANAQVVMICGKSKELK---RSLTAKFKLTRMYLILGYTKHMNEWMASSQ 275
           : |        ||::||| :|:|    ||:  |  :     : |:   | | | :
Db     364 skalydenhgepigqvlvicghnkklagrlrsidwkvpvq----vkgfvtkmeecmgacd 419

Qy     276 LMITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASL 335
           :||| |  || |   | :|:|  :   |||  |  |  | | || : :|:|   |||

Db     420 ciitkagpgtiaeamirglpiilndyiagqeagnvpyvvengcgkfskspkeianivakw 479

Qy     336 TNGNEQLTNMISTMEQDKIKYATQ----TICRDLLDLIGHSSQPQEIYGKVPLYA 386
                 : :  | |: :: |       |  ||.:|:   |          || |:
Db     480 fgpk---adellimsqnalrlarpdavfkivhdlhelvkqrs-------fvpqys 524


RESULT   6
AAB19039
ID   AAB19039 standard; protein; 522 AA.
XX
AC   AAB19039;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Amino acid sequence of monogalactosyl-diacylglycerol (MGDG) synthase.
XX
KW   Monogalactosyl-diacylglycerol synthase; MGDG synthase; Toxoplasma;
KW   apicomplex parasite; herbicide; antiparasitic; Plasmodium; Eimeria;
KW   acquired immune deficiency syndrome; coccidiosis.
XX
OS   Spinacia oleracea.
XX
PN   WO200056919-A1.
XX
PD   28-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-FR00658.
XX
PR   19-MAR-1999;   99FR-0003434.
XX
PA   (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI   Marechal E,  Block M,  Joyard J,  Douce R;
XX
DR   WPI; 2000-602227/57.
XX
PT   Use of monogalactosyl diacylglycerol synthase for identifying its
PT   specific inhibitors, potentially useful as antiparasitic agents and
PT   herbicides -
XX
PS   Example 2; Fig 2; 33pp; French.
XX
CC   The present sequence represents a monogalactosyl-diacylglycerol (MGDG)
CC   synthase. MGDG is present in all plastids tested and is essential for
CC   cell survival, but is not present in other membrane structures, or in
CC   animal cells, and so represents a specific target. MGDG synthase or a
CC   plastid membrane isolated from plants is used to select and screen for
CC   specific inhibitors of MGDG synthase. These inhibitors are suitable as
CC   active agents against apicomplex parasites or as herbicides. The
CC   inhibitors are used as antiparasitic agents, especially against
CC   Plasmodium, Toxoplasma (particularly in patients with acquired immune
CC   deficiency syndrome) or Eimeria (coccidiosis in cattle or poultry), and
CC   as herbicides.
XX
SQ   Sequence   522 AA;

  Query Match            13.6%;  Score 279.5;  DB 21;  Length 522;
  Best Local Similarity  24.5%;  Pred. No. 5.5e-18;
  Matches   96;  Conservative   64;  Mismatches  191;  Indels   41;  Gaps   11;

Qy       6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSF 65
           |::||:    | ||    ::|    |:   |   |   ||: |  |        |  |
Db     131 krvlilmsdtggghrasaeaikaafneefgddyqvfvtdlwsehtp--------wpfnql 182

Qy      66 KYFRN-------MYKGFYYSRPDKL----DKCFYKYYGLNKLINLLIKEKPDLILLTFP- 113
             |       ::|  ||    ::    :      :   ::   |:| :||:|:   |
Db     183 prsynflvkhgplwkmmyygtsprvihqsnfaatsvfiarevarglmkyqpdiiisvhpl 242

Qy     114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
               |:  :        |   ||:||    |  |     || |  : |  :     |: |
Db     243 mqhvplrilrgrgllekivfttvvtdlstchptwfhklvtrcycpsnevakratkaglqp 302
```

```
Qy    170 STVKVTGIPIDNKFETPINQKQWL-IDNNLDPDKQTILMSAGAFGV------SKGFDTMI 222
          | :|| |:|:   |   :  |  |  :  :|     :|:  |  |:       ::     :
Db    303 sqikvyglpvrpsfvrsvrpknelrkelgmdehlpavllmgggegmgpieataralgnal 362

Qy    223 TDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLMITKP 281
           |        |:::|||::|:|   |::  :|:     : |:   : | | :   :|||
Db    363 ydanlgeptgqllvicgrnkklagklssidwkip--vqvkgfvtkieecmgacdciitka 420

Qy    282 GGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTNGNEQ 341
          |  || |   | :|:|  :   |||  |  |  | | ||   :|:|  | |:
Db    421 gpgtiaeamirglpiilndyiagqeagnvpyviengigkylkspkeiaktvsqwfgpk-- 478

Qy    342 LTNMISTMEQDKIKYATQ----TICRDLLDLI 369
            | :  | |: :|:|       |  || :|:
Db    479 -anelqimsqnalkharpdavfkivhdldelv 509


RESULT   7
AAG42414
ID   AAG42414 standard; Protein; 492 AA.
XX
AC   AAG42414;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 52894.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
```

```
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             13.3%;  Score 272;  DB 21;  Length 492;
  Best Local Similarity   24.9%;  Pred. No. 2.6e-17;
  Matches  104;  Conservative   71;  Mismatches  184;  Indels   58;  Gaps   16;

Qy      6 KKILIITGSFGNGHMQVTQSIVNQLN-DMNLDHLSVIEHDLFME---------AHPILTS 55
          ||:||:     | ||     ::|     | :    :: | :|    |:|          | :|
Db     83 kkvlilmsdtggghrasaeairaafnqefgdeyqsgle---flelekigfrclvhvfitd 139

Qy     56 IC---KKWYINSFKYFRN-------MYKGFYYSRPDKL----DKCFYKYYGLNKLINLLI 101
          :        |  |       |         ::| ||      ::      :      :   ::  |:
Db    140 lwtdhtpwpfnqlprsynflvkhgtlwkmtyygtsprivhqsnfaatstfiareiaqglm 199

Qy    102 KEKPDLILLTFP----TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATK 156
          | :||:|:    |       |:  :  ::       |     ||:||      |     |     || |   :
Db    200 kyqpdiiisvhplmqhvplrvlrskgllkkivfttvitdlstchptwfhklvtrcycpst 259

Qy    157 ETKQDFIDVGIDPSTVKVTGIPIDNKFETPINQKQWL-----IDNNLDPDKQTILMSAGA 211
          |  :     |:: | :|| |:|:    |   |:   |   |        :| ||        :|:  |
Db    260 evakraqkagletsqikvyglpvrpsfvkpvrpkvelrrelgmdenl----pavllmggg 315

Qy    212 FGV------SKGFDTMITDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYT 264
           |:       ::          :  |          |    ||::|||::|:|:   |::   :|:        :  |:
Db    316 egmgpieataraladalydknlgeavgqvliicgrnkklqsklssldwkip--vqvkgfi 373

Qy    265 KHMNEWMASSQLMITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADT 324
           |  |  |  :     :|||  |    ||  |      |  :|:|          |||   |    |   |  |  ||  :  :
Db    374 tkmeecmgacdciitkagpgtiaeamirglpiilngyiagqeagnvpyvvengcgkfsks 433

Qy    325 PEEAIKIVASLTNGNEQLTNMISTMEQDKIKYATQ----TICRDLLDLI-GHSSQPQ 376
          |:|   ||||            :   :   |  |: :: |           |  |:  :|:     :|  ||
Db    434 pkeiskivadwfg---paskeleimsqnalrlakpeavfkivhdmhelvrkknslpq 487


RESULT   8
AAG42413
ID    AAG42413 standard; Protein; 551 AA.
XX
AC    AAG42413;
XX
DT    18-OCT-2000  (first entry)
XX
DE    Arabidopsis thaliana protein fragment SEQ ID NO: 52893.
XX
KW    Protein identification; signal transduction pathway; metabolic pathway;
KW    hybridisation assay; genetic mapping; gene expression control; promoter;
KW    termination sequence.
XX
OS    Arabidopsis thaliana.
XX
PN    EP1033405-A2.
XX
PD    06-SEP-2000.
XX
PF    25-FEB-2000; 2000EP-0301439.
XX
PR    25-FEB-1999;   99US-0121825.
PR    05-MAR-1999;   99US-0123180.
PR    09-MAR-1999;   99US-0123548.
PR    23-MAR-1999;   99US-0125788.
PR    25-MAR-1999;   99US-0126264.
PR    29-MAR-1999;   99US-0126785.
PR    01-APR-1999;   99US-0127462.
PR    06-APR-1999;   99US-0128234.
PR    08-APR-1999;   99US-0128714.
PR    16-APR-1999;   99US-0129845.
PR    19-APR-1999;   99US-0130077.
PR    21-APR-1999;   99US-0130449.
PR    23-APR-1999;   99US-0130510.
PR    23-APR-1999;   99US-0130891.
PR    28-APR-1999;   99US-0131449.
PR    30-APR-1999;   99US-0132048.
```

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

```
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             13.3%;  Score 272;  DB 21;  Length 551;
  Best Local Similarity   24.9%;  Pred. No. 3.1e-17;
  Matches  104;  Conservative   71;  Mismatches  184;  Indels   58;  Gaps   16;

Qy       6 KKILIITGSFGNGHMQVTQSIVNQLN-DMNLDHLSVIEHDLFME---------AHPILTS 55
           ||:||:    | ||    ::|    | :   :: | :|   |:|           |  :|
Db     142 kkvliilmsdtggghrasaeairaafnqefgdeyqsgle---flelekigfrclvhvfitd 198

Qy      56 IC---KKWYINSFKYFRN-------MYKGFYYSRPDKL----DKCFYKYYGLNKLINLLI 101
           :      | |       |        ::|  ||    ::    :      :   ::   |:
Db     199 lwtdhtpwpfnqlprsynflvkhgtlwkmtyygtsprivhqsnfaatstfiareiaqglm 258

Qy     102 KEKPDLILLTFP----TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATK 156
           | :||:|:   |     |:  : ::     |   ||:||    |  |     || |  :
Db     259 kyqpdiiisvhplmqhvplrvlrskgllkkivfttvitdlstchptwfhklvtrcycpst 318

Qy     157 ETKQDFIDVGIDPSTVKVTGIPIDNKFETPINQKQWL-----IDNNLDPDKQTILMSAGA 211
           |  :     |:: | :|| |:|:   |  |:  |  |     :| ||      :|:  |
Db     319 evakraqkagletsqikvyglpvrpsfvkpvrpkvelrrelgmdenl----pavllmggg 374

Qy     212 FGV------SKGFDTMITDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYT 264
            |:      ::     : |     |  ||::|||::|:|:  |::  :|:     : |:
Db     375 egmgpieataraladalydknlgeavgqvliicgrnkklqsklssldwkip--vqvkgfi 432

Qy     265 KHMNEWMASSQLMITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADT 324
           | | | :   :|||  |  || |   | :|:|     |||  |  |  | | || : :
Db     433 tkmeecmgacdciitkagpgtiaeamirglpiilngyiagqeagnvpyvvengcgkfsks 492

Qy     325 PEEAIKIVASLTNGNEQLTNMISTMEQDKIKYATQ----TICRDLLDLI-GHSSQPQ 376
           |:|  ||||        :  :  | |: :: |       |  |: :|:   :| ||
Db     493 pkeiskivadwfg---paskeleimsqnalrlakpeavfkivhdmhelvrkknslpq 546
```

```
RESULT   9
AAG42415
ID   AAG42415 standard; Protein; 404 AA.
XX
AC   AAG42415;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 52895.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
```

```
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
```

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.3%; Score 252; DB 21; Length 404;
Best Local Similarity 24.6%; Pred. No. 1.5e-15;
Matches 100; Conservative 69; Mismatches 180; Indels 58; Gaps 16;

```
Qy     16 GNGHMQVTQSIVNQLN-DMNLDHLSVIEHDLFME---------AHPILTSIC---KKWYI 62
          | ||    ::|    | :   :: | :|   |:|           |  :| :      |
Db      5 ggghrasaeairaafnqefgdeyqsgle---flelekigfrclvhvfitdlwtdhtpwpf 61

Qy     63 NSFKYFRN-------MYKGFYYSRPDKL----DKCFYKYYGLNKLINLLIKEKPDLILLT 111
          |      |       ::|  ||    ::    :      :   ::   |:| :||:|:
Db     62 nqlprsynflvkhgtlwkmtyygtsprivhqsnfaatstfiareiaqglmkyqpdiiisv 121

Qy    112 FP----TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATKETKQDFIDVG 166
          |     |:  : ::     |   ||:||    |  |     || |  : |  :     |
Db    122 hplmqhvplrvlrskgllkkivfttvitdlstchptwfhklvtrcycpstevakraqkag 181

Qy    167 IDPSTVKVTGIPIDNKFETPINQKQWL-----IDNNLDPDKQTILMSAGAFGV------S 215
          :: | :|| |:|:   |  |:  | .|     :| ||      :|:  |  |:      :
Db    182 letsqikvyglpvrpsfvkpvrpkvelrrelgmdenl----pavllmgggegmgpieata 237

Qy    216 KGFDTMITDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASS 274
          :    : |     |  ||::|||::|:|:  |::  :|:     : |:   | | | :
Db    238 raladalydknlgeavgqvliicgrnkklqsklssldwkip--vqvkgfitkmeecmgac 295

Qy    275 QLMITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVAS 334
          :||| |  || |   | :|:|     |||  |  |  | | || : :|:|  ||||
Db    296 dciitkagpgtiaeamirglpiilngyiagqeagnvpyvvengcgkfskspkeiskivad 355

Qy    335 LTNGNEQLTNMISTMEQDKIKYATQ----TICRDLLDLI-GHSSQPQ 376
               :  : | |: :: |       |  |: :|:   :| ||
Db    356 wfg---paskeleimsqnalrlakpeavfkivhdmhelvrkknslpq 399
```

RESULT 10
AAW88310
ID AAW88310 standard; Protein; 374 AA.
XX

```
AC   AAW88310;
XX
DT   26-APR-1999  (first entry)
XX
DE   E. coli colitose or glucose transferase.
XX
KW   O antigen; O111 antigen; wbdM gene; colitose transferase;
KW   glucose transferase; diarrhoea; haemorrhagic colitis; diagnosis.
XX
OS   Escherichia coli.
XX
PN   WO9850531-A1.
XX
PD   12-NOV-1998.
XX
PF   01-MAY-1998;   98WO-AU00315.
XX
PR   22-JUL-1997;   97AU-0008162.
PR   01-MAY-1997;   97AU-0006545.
XX
PA   (UNSY ) UNIV SYDNEY.
XX
PI   Reeves PR,  Wang L;
XX
DR   WPI; 1999-059669/05.
DR   N-PSDB; AAX06748.
XX
PT   Nucleic acid molecules specific for bacterial polysaccharide
PT   antigens - useful for detecting specific strains in, e.g. food,
PT   faeces or patient samples
XX
PS   Disclosure; Fig 7; 165pp; English.
XX
CC   This is the amino acid sequence of the protein encoded by the wbdM
CC   gene of a gene cluster (see AAX06748) involved in the biosynthesis of
CC   the Escherichia coli O111 O antigen.  The protein shows high
CC   homology with TrsE of Yersinia enterocholitica, and is predicted to
CC   be a colitose or glucose transferase.  The use of nucleic acid
CC   molecules derived from particular assembly and transport genes,
CC   particularly wbd (transferase), wzx (flippase) and wzy (polymerase)
CC   genes, within O antigen gene clusters improves the specificity of
CC   methods for the detection and identification of O antigens, e.g. in
CC   testing food- or faecal-derived samples, or samples from patients.
CC   The O antigen is a major virulence factor of enteropathogenic E.
CC   coli strains that cause diarrhoea and haemorrhagic colitis.
XX
SQ   Sequence   374 AA;
```

```
  Query Match             6.8%;  Score 139;  DB 20;  Length 374;
  Best Local Similarity   20.9%;  Pred. No. 7.3e-05;
  Matches   84;  Conservative   69;  Mismatches  163;  Indels   86;  Gaps   18;

Qy       6 KKILIITGSFGNGHMQVTQSIVNQL----NDMNLDHLSVIEHDLFMEAHPILTSICKKWY 61
           | : ||||    |   :   : :|:    :|:|:  |: |      |  |  |      |
Db       2 kivyiitgltcggaehlmtqladqmfirghdvniicltgis-----evkp--tqninihy 54

Qy      62 INSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDLI-----------L 109
           :|  | ||: ::  :                ::  :::  |||:|
Db      55 vnmdknfrsffralf----------------qvkkiivalkpdiihshmfhanifsrf 96

Qy     110 LTFPTPVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
           :    | : ::    | |      |  |||  :::   :|     :||  |:||     |
Db      97 irmlipavplictahnkneggnarmfcyrl-sdflasittn---vskeavqefiarkatp 152

Qy     170 STVKVTGIP---IDNKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITD-- 224
             |: ||     |||:  || ::   |     |   :|:: |    :| :  ::
Db     153 kn-kiveipnfintnkfdfdinvrkktrdafnlkdstavllavgrlveakdypnllnain 211

Qy     225 --ILAKSANAQ--VVMICG----KSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQL 276
             ||:|::|    :::| |    ::| |   |  :  |      ||    : | | :: |
Db     212 hlilsktsncndfilliagdgalrnkll--dlvcqlnlvdkvfflgqrsdikelmcaadl 269
```

```
Qy     277 MITKPG----GITITEGFARCIPMIFLNPAPGQE----------LENAFYFEEKGFGKIA 322
           :        |: : |  |   |::  :    :|          : |     |    |||
Db     270 fvlssewegfglvvaeamacerpvvatdsggvkevvgphndvipvsnhillae----kia 325

Qy     323 DT---PEEAIKIVASLTN----GNEQLTNMISTMEQDKIKYA 357
           :|    : | ||:          |  :  ::|  |:   ||:
Db     326 etlkiddnarkiigmknreyivsnfsiktivsewerlyfkys 367
```

```
RESULT  11
AAW51348
ID   AAW51348 standard; Protein; 352 AA.
XX
AC   AAW51348;
XX
DT   03-SEP-1998  (first entry)
XX
DE   Streptococcus pneumoniae stem peptide biosynthetic murG protein.
XX
KW   Streptococcus pneumoniae; stem peptide biosynthesis; murG; inhibitor;
KW   infection; antibiotic resistant; microorganism; bacteria;
KW   cell wall biosynthesis.
XX
OS   Streptococcus pneumoniae.
XX
PN   EP843012-A2.
XX
PD   20-MAY-1998.
XX
PF   18-NOV-1997;   97EP-0309247.
XX
PR   19-NOV-1996;   96US-0751474.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Hoskins JA,  Peery RB,  Skatrud PL;
XX
DR   WPI; 1998-263289/24.
DR   N-PSDB; AAV07310.
XX
PT   Recombinant Streptococcus pneumoniae MurG protein encoded by nucleic
PT   acid expressed in host cells - useful in screening for inhibitors of
PT   MurG protein and stem peptide biosynthesis
XX
PS   Claim 1; Page 15-16; 20pp; English.
XX
CC   The present sequence represents the stem peptide biosynthetic murG
CC   protein from Streptococcus pneumoniae. The murG protein can be used in
CC   the development of drugs for treating infections by antibiotic-resistant
CC   microorganisms. MurG inhibitory compounds are used to identify
CC   inhibitors which function by interfering with the stem peptide pathway
CC   in bacterial cell wall biosynthesis rather than with the peptidoglycan
CC   crosslinking step, which are useful for treating infections of
CC   beta-lactam resistant bacteria.
XX
SQ   Sequence   352 AA;
```

```
  Query Match             5.8%;  Score 119;  DB 19;  Length 352;
  Best Local Similarity   20.4%;  Pred. No. 0.0053;
  Matches   79;  Conservative   80;  Mismatches  155;  Indels   74;  Gaps   19;

Qy       7 KILIITGSFGNGHMQVTQSI----------VNQLND-MNLDHLSVIEHDLFMEAHPILTS 55
           | :: ||    ||: :   :          |: : |   ::|  :::  | :  | | |
Db       2 kkivftgggtvghvtlnlllmpkfiedgwevhyigdkrgiehqeilksgldvtfhsiatg 61

Qy      56 ICKKWYINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDLI-----LL 110
             ::                |:|  : ||  |   :|: : : :::: :|  :      :
Db      62 klrr---------------yfswqnmld-vfkvgwgivqslfimlrlrpqtlfskggfv 104

Qy     111 TFPTPVMSVLTEQFNINIPVATVMTDYR--LHKNWITPYSTRYYVATKETKQDFIDVGID 168
           : | ||::         :||    :|    |       ::|: | :| |       |
Db     105 svp-pviaarvs----gvpvfihesdlsmglankiaykfatkmy-stfeqasslskvehv 158
```

```
Qy    169 PSTVKVTGIPIDNKFETPINQKQWLID--NNLDPDKQTILMSAGAFGVSKGFDTMITDIL 226
          :  ||:    | |   |      |:|   : :    |:|   |: | :: |: ::||
Db    159 gavtkvs----dqknpep----delvdiqthfnhklptvlfvggsag-arvfnqlvtdhk 209

Qy    227 AK-SANAQVVMICGKS--KELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGG 283
          : :    :: : | |   || ::|     :| :|  |          |  : :::|: |
Db    210 kelteryniinltgdsslnelsqnlfrvdyvtdlyqpl---------meladvvvtrgga 260

Qy    284 ITITEGFA----RCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTNGN 339
          || |  |      |  :    : | ::||| || :||: :  |  |  :    :| :
Db    261 ntifellaiaklhvivplgreasrgdqienaayfvkkgyae--dlqesdl----tldsle 314

Qy    340 EQLTNMISTMEQDKIKYATQTICRDLLD 367
          |:|::::|  |  : |       : | |
Db    315 eklshllshkedyqakmkaskelkslad 342


RESULT 12
AAR26154
ID   AAR26154 standard; Protein; 534 AA.
XX
AC   AAR26154;
XX
DT   27-JAN-1993  (first entry)
XX
DE   HUG-Br2.
XX
KW   Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW   monoglucuronide; diglucuronide.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          12..22
FT                   /note= "putative membrane-insertion signal"
FT   Region          492..508
FT                   /note= "putative membrane-anchoring peptide"
FT   Modified-site   348
FT                   /note= "predicted Asn-linked glycosylation site"
FT   Misc-difference 282..285
FT                   /note= "residues encoded by TGCCAACGGGAAG !"
XX
PN   WO9212987-A.
XX
PD   06-AUG-1992.
XX
PF   10-JAN-1992;   92WO-US00282.
XX
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS,  Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   N-PSDB; AAQ27369.
XX
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 9A-I; 99pp; English.
XX
CC   Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC   been isolated.  They are referred to as HUGBr1 (AAQ27369) and HUGBr2
CC   (AAQ27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC   upon expression individually in COS-1 cells, encode isoforms that
CC   catalyse the formation of the two bilirubin monoglucuronides and
CC   the diglucuronide.
CC   The cDNAs contain identical 3' ends (1469 bp in length) to each
CC   other and to that of the human phenol transferase cDNA, HLUGP1
CC   (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC   In contrast, they have unique 5' ends.
XX
SQ   Sequence   534 AA;


  Query Match             5.3%;  Score 109.5;  DB 13;  Length 534;
  Best Local Similarity   20.4%;  Pred. No. 0.078;
  Matches   58;  Conservative   54;  Mismatches  108;  Indels   65;  Gaps   12;

Qy     93 LNKLINLLIKEKPDLILLTFPTPVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYY 152
          | :: |:|       |  ||  |  |: :| |   : |  :::   :   |:
Db    208 lqrvknmlyplalsyichtfsapyaslaselfqrevtvqdllssasv---wl-------- 256

Qy    153 VATKETKQDFI---DVGIDPSTVKVTGI------PIDNKFETPINQKQWLIDNNLDPDKQ 203
                : ||:      | |: | : ||      |:  :||  ||
Db    257 -----frsdfvkdyprpimpnmvfiggincangkplsqefeayin-------------- 296

Qy    204 TILMSAGAFG-VSKGFDTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLTRMYLILG 262
              ::|  | |    ::|:::|  | | |    |     :: :::  ::  ||   :
Db    297 ----asgehgivvfslesmvseipekkama----iadalgkipqtvlwrytgtrpsnlan 348

Qy    263 YTKHMNEWMASSQLM--------ITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFE 314
          |  : :|:  : |:         ||  |   : |     :||: : |  | :::||   |
Db    349 nt-ilvkwlpqndllghpmtrafithagshgvyesicngvpmvmm-plfgdqmdnakrme 406

Qy    315 EKGFGKIAD----TPEEAIKIVASLTNGNEQLTNM--ISTMEQDK 353
          || |   :    | |:      :: |     |:  :|:: :|:
Db    407 tkgagvtlnvlemtsedlenaqkavindksykenimrlsslhkdr 451


RESULT 13
AAY78935
ID   AAY78935 standard; Protein; 530 AA.
XX
AC   AAY78935;
XX
DT   05-JUN-2000  (first entry)
XX
DE   Human UDP-glucuronosyltransferase 2B15 amino acid sequence.
XX
KW   UDP-glucuronosyltransferase 2B15; UGT2B15; polymorphism; metabolism;
KW   drug interaction; detect; human; single nucleotide polymorphism; SNPs.
XX
OS   Homo sapiens.
XX
PN   WO200006776-A1.
XX
PD   10-FEB-2000.
XX
PF   22-JUL-1999;   99WO-US16675.
XX
PR   28-JUL-1998;   98US-0094391.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Galvin M,  Miller A,  Penny L,  Riedy M;
XX
DR   WPI; 2000-195321/17.
DR   N-PSDB; AAZ95206.
XX
PT   Novel human UDP-glucuronosyltransferase sequence, polymorphisms for
PT   genotyping individuals to predict rate of metabolism of substrates and
PT   for identifying potential drug interactions  -
XX
PS   Disclosure; Page 59-60; 72pp; English.
XX
CC   This sequence represents the human UDP-glucuronosyltransferase 2B15
CC   (UGT2B15) amino acid sequence. UDP-glucuronosyltransferase (UGTs) are a
CC   family of enzymes that catalyse the glucuronic acid conjugation of a
CC   wide range of endogenous and exogenous substrates. The UGT2B gene
CC   subfamily encode steroid metabolizing isoforms in the liver. Alteration
CC   of the expression or function of UGTs may effect drug metabolism. The
CC   invention relates to non-chromosomal nucleic acid molecules, which
```

```
CC   comprise human UGT2B sequence polymorphisms (see AAZ95051-Z95110). Probes
CC   which detect the UGT2B locus polymorphisms can be used to detect altered
CC   UGT2B metabolism of a substrate in an individual. The nucleic acid
CC   molecules comprising a human UGT2B sequence polymorphism can be used in
CC   screening assays for genotyping individuals, also to predict their rate
CC   of metabolism of UGT2B substrate, potential drug-drug interactions and
CC   adverse side effects. The polymorphisms can be used as single nucleotide
CC   polymorphisms (SNPs) for detecting genetic linkage related to phenotypic
CC   variation in activity or expression of UGT2B protein. The polymorphism
CC   containing nucleic acid molecules may also be used for generating
CC   genetically modified non-human animals and for obtaining site specific
CC   gene modification in cell lines.
XX
SQ   Sequence   530 AA;


  Query Match             5.2%;  Score 106;  DB 21;  Length 530;
  Best Local Similarity   19.2%;  Pred. No. 0.16;
  Matches   93;  Conservative   78;  Mismatches  160;  Indels  154;  Gaps   24;

Qy       1 MVTQNKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKW 60
           :| :  :: ::| |          ::||        | |:    :| :|  ||: |
Db      46 lvqrghevtvltss--------astlvnasks------saik----levyp--tsltknd 85

Qy      61 YINSF-----KYFRNMYKGFYYSRPDKLDKCFYKYYG----------LNKLINLLIKE-K 104
             :|      ::   : |  ::|   :| :  ::||           ||| : : ::| |
Db      86 ledsllkildrwiygvsknfwsyfsqlqelcweyydysnklckdavlnkklmmklqesk 145

Qy     105 PDLILLTFPTPVMSVLTEQFNI-------------------------NIPVA-TVMTDY 137
           |:||     |   :| | |||                           :||  : ::|
Db     146 fdviladalnpcgellaelfnipflyslrfsvgytfeknggglflppsyvpvvmselsdq 205

Qy     138 RLH----KN---------WITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGIPIDNKFE 184
            :     ||         |   |        |:  | : :|   |:|:           ||
Db     206 mifmeriknmihmlyfdfwfqiyd------lkkwdqfysevlgrpttl----------fe 249

Qy     185 TPINQKQWLIDN-----------NLD--------------PDKQTILMSAGAFGV---S 215
           |    : |||            |:|              : :  : |:|  |:   |
Db     250 tmgkaemwlirtywdfefprpflpnvdfvgglhckpakplpkemeefvqssgengivvfs 309

Qy     216 KGFDTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQ 275
           |  :||:::  :|||    ||     :: : :  :|   :    ||    : :|:  :
Db     310 lg--smisnmseesan----miasalaqipqkvlwrfd-gkkpntlgsntrlykwlpqnd 362

Qy     276 LM--------ITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEE 327
           |:        ||  |   | |     |||: : |    : :|  : : ||
Db     363 llghpktkafithggtngiyeaiyhgipmvgi-plfadqhdniahmkakg---------- 411

Qy     328 AIKIVASLTNGNEQLTNMISTMEQDKI-KYATQTICRDLLDLIGHSSQPQEIYGKVPLYA 386
           |   |   |  :  | | : ::  | : |     : |       |  || :   :   :
Db     412 aalsvdirtmssrdllnalksvindpvykenvmklsri------hhdqpmkpldravfwi 465

Qy     387 RFFVK 391
           | ::
Db     466 efvmr 470


RESULT  14
AAR76063
ID   AAR76063 standard; Protein; 934 AA.
XX
AC   AAR76063;
XX
DT   15-JAN-1996  (first entry)
XX
DE   Human MSH2 protein, homologue of the E.coli mutS gene product.
XX
KW   Mismatch repair; MSH2; primer; identification; defect; alteration;
KW   cancer; tumour; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO9514085-A2.
XX
PD   26-MAY-1995.
XX
PF   17-NOV-1994;   94WO-US13385.
XX
PR   13-JUN-1994;   94US-0259310.
PR   17-NOV-1993;   93US-0154792.
PR   07-DEC-1993;   93US-0163449.
XX
PA   (UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
PA   (DAND ) DANA FARBER CANCER INST.
XX
PI   Fishel R,  Kolodner RD,  Reenan RAG;
XX
DR   WPI; 1995-200377/26.
DR   N-PSDB; AAQ93901.
XX
PT   Determining alteration in human mismatch repair pathways - used in
PT   the diagnosis, prognosis and therapy of cancers and in screening
PT   assays
XX
PS   Claim 27; Page 159-163; 256pp; English.
XX
CC   AAQ93901 is the human mismatch repair pathway gene MSH2. Defects or
CC   alterations in such a gene result in the accumulation of unstable
CC   repeated DNA sequences, a feature of a number of different cancers.
CC   The identification of a defect in the mismatch repair pathway can
CC   be diagnostic of a predisposition to cancer and prognostic for a
CC   particular mammalian cancer e.g colorectal, ovarian, endometrial
CC   (uterine), renal, bladder, skin, rectal and bowel. The nucleotide
CC   sequences and polypeptides of the hMSH2 gene may also be used for
CC   therapy and in vaccines.
XX
SQ   Sequence   934 AA;


  Query Match             5.0%;  Score 102.5;  DB 16;  Length 934;
  Best Local Similarity   22.3%;  Pred. No. 0.81;
  Matches   77;  Conservative   45;  Mismatches  131;  Indels   93;  Gaps   17;

Qy      18 GHMQVTQSIVNQLNDMNL--DHLSVIE----------HDLFMEAHPILTSICKKWYI 62
           |   | | |   | | |   : |:::|             || :   | |  : ||
Db     338 gqrlvnqwikqplmdknrieerlnlveafvedaelrqtlqedl-lrrfpdlnrlakk--- 393

Qy      63 NSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLL-IKEKPD-----LILLTFPTPV 116
                     :  :   |  |:  | |:|:| |::   || :     |:|  | ||:
Db     394 ------------fqrqaanlqdcyrlyqginqlpnviqalekhegkhqklllavfvtpl 440

Qy     117 MSVLTE--QFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKV 174
             : ::  :|   |      |     |:  : |                    ||:  ::
Db     441 tdlrsdfskfqemiettldmdqvenheflvkp------------------sfdpnlsel 481

Qy     175 TGI--PIDNKFE-TPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSAN 231
             |   :: | : | |:  :   |  ||| ||  | |:  ||           :|  : |
Db     482 reimndlekkmqstlisaar---dlgldpgkqikldssaqfgyyfrvtckeekvlrnnkn 538

Qy     232 AQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITITEGFA 291
             | |     :   |     |||   |   |||:  |:  :   :: :   : |: |:
Db     539 fstvdiqkngvkftns-----klts--lneeytknkteyeeaqdaivkei--vnissgyv 589

Qy     292 RCIPMIFLNPAPGQELENAFYFE----------------EKGFGKI 321
           ||  ||    | |:    |                  ||| |:|
Db     590 e--pmqtlndvlaq-ldavvsfahvsngapvpyvrpailekgqgri 632


RESULT  15
AAR75411
ID   AAR75411 standard; Protein; 934 AA.
XX
AC   AAR75411;
XX
DT   14-NOV-1995  (first entry)
XX
```

DE Human MSH2.
XX
KW MSH2; mutator gene; DNA mismatch repair;
KW hereditary non-polyposis colorectal cancer.
XX
OS Homo sapiens.
XX
PN WO9515381-A.
XX
PD 08-JUN-1995.
XX
PF 02-DEC-1994; 94WO-US13805.
XX
PR 02-DEC-1993; 93US-0160295.
XX
PA (UYJO) UNIV JOHNS HOPKINS.
XX
PI De La Chapelle A, Kinzler KW, Vogelstein B;
XX
DR WPI; 1995-215261/28.
DR N-PSDB; AAQ87269.
XX
PT New cDNA from human MSH2 gene encoding DNA mismatch repair enzyme -
PT its mutants causing hereditary non-polyposis colorectal cancer and
PT derived proteins useful in cancer prevention or treatment and for
PT diagnosis and screening.
XX
PS Disclosure; Page 37-38; 60pp; English.
XX
CC cDNA from human colon cancer cells was amplified using primers
CC previously used to isolate the yeast MSH2 gene from homology with
CC MutS. The insert in isolated clone pNP-23 was used to screen
CC cDNA libraries, and positive clones were used in a chromosome
CC walking procedure to identify the entire coding sequence (given
CC in AAQ87269) of the human MSH2 gene.
XX
SQ Sequence 934 AA;

Query Match 5.0%; Score 102.5; DB 16; Length 934;
Best Local Similarity 22.3%; Pred. No. 0.81;
Matches 77; Conservative 45; Mismatches 131; Indels 93; Gaps 17;

```
Qy     18 GHMQVTQSIVNQLNDMNL--DHLSVIE-------------HDLFMEAHPILTSICKKWYI 62
Db    338 gqrlvnqwikqplmdknrieerlnlveaivedaelrqtlqedl-lrrfpdlnrlakk--- 393

Qy     63 NSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLL-IKEKPD-----LILLTFPTPV 116
Db    394 ------------fqrqaanlqdcyrlyqginqlpnviqalekhegkhqklllavfvtpl 440

Qy    117 MSVLTE--QFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKV 174
Db    441 tdlrsdfskfqemiettldmdqvenheflvkp------------------sfdpnlsel 481

Qy    175 TGI--PIDNKFE-TPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSAN 231
Db    482 reimndlekkmqstlisaar---dlgldpgkqikldssaqfgyyfrvtckeekvlrnnkn 538

Qy    232 AQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITITEGFA 291
Db    539 fstvdiqkngvkftns-----klts--lneeytknkteyeeaqdaivkei--vnissgyv 589

Qy    292 RCIPMIFLNPAPGQELENAFYFE----------------EKGFGKI 321
Db    590 e--pmqtlndvlaq-ldavvsfahvsngapvpyvrpailekgqgri 632
```

Search completed: June 29, 2001, 08:59:04
Job time: 177 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 29, 2001, 09:00:19 ; Search time 28.15 Seconds
(without alignments)
1058.055 Million cell updates/sec

Title: US-09-668-788-4
Perfect score: 2048
Sequence: 1 MVTQNKKILIITGSFGNGHM.....SSQPQEIYGKVPLYARFFVK 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	591.5	28.9	382	2	C69935	cell wall synthesi
2	309	15.1	464	2	C84499	probable monogalac
3	299	14.6	411	2	F75439	probable cell wall
4	296.5	14.5	468	2	T52269	1,2-diacylglycerol
5	284	13.9	533	2	T05092	probable 1,2-diacy
6	283	13.8	525	2	T10478	probable 1,2-diacy
7	245	12.0	373	2	B69860	conserved hypothet
8	230	11.2	374	2	E83894	hypothetical prote
9	136	6.6	363	2	F70195	UDP-N-acetylglucos
10	127.5	6.2	1157	2	S38160	NUP133 protein - y
11	124	6.1	363	2	E83970	UDP-N-acetylglucos
12	123.5	6.0	344	2	C70401	phospho-N-acetylmu
13	122	6.0	447	2	S35481	SRPM54 protein - M
14	121	5.9	359	2	T44332	hypothetical prote
15	117.5	5.7	353	2	C64661	transferase, pepti
16	116.5	5.7	353	2	G71852	udp-n-acetylglucos
17	116.5	5.7	357	2	E86823	peptidoglycan synt
18	115.5	5.6	398	2	F64456	hypothetical prote
19	113.5	5.5	765	2	S70962	regulatory protein
20	113	5.5	950	2	A71655	hypothetical prote
21	112.5	5.5	383	2	E70156	lipopolysaccharide
22	111.5	5.4	363	2	JC1275	phospho-N-acetylmu
23	109.5	5.3	531	2	A35343	glucuronosyltransf
24	109	5.3	354	2	D84955	hypothetical prote
25	108.5	5.3	363	1	C72590	probable hexosyltr
26	108.5	5.3	495	2	T28722	hypothetical prote
27	106	5.2	274	2	S60880	phosphorylation-ac
28	106	5.2	339	2	B72402	UDP-N-acetylglucos
29	106	5.2	390	2	A86065	hypothetical prote
30	106	5.2	530	2	A48633	glucuronosyltransf
31	106	5.2	533	2	T27589	hypothetical prote
32	106	5.2	1082	2	T43990	hypothetical prote
33	104.5	5.1	529	2	A24600	glucuronosyltransf
34	104	5.1	1196	2	JQ1467	toxin, nontoxic co
35	104	5.1	1196	2	S46430	botulinum neurotox
36	103.5	5.1	484	2	E64432	spore coat polysac
37	102.5	5.0	534	2	T19944	hypothetical prote
38	102.5	5.0	934	2	I64819	DNA mismatch repai
39	102	5.0	389	2	E65182	bacteriophage N4 a
40	102	5.0	398	2	F40511	hypothetical prote
41	101.5	5.0	368	2	F70346	mannosyltransferas
42	101.5	5.0	474	2	A64691	type III restricti
43	101.5	5.0	535	2	I57961	glucuronosyltransf
44	101.5	5.0	1248	2	A47445	reverse gyrase - S
45	101	4.9	530	2	T19365	hypothetical prote

ALIGNMENTS

RESULT 1
C69935
cell wall synthesis homolog ypfP - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: C69935
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033
A;Accession: C69935
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-382 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14110.1; PID:e11836
A;Experimental source: strain 168
C;Genetics:
A;Gene: ypfP

Query Match 28.9%; Score 591.5; DB 2; Length 382;
Best Local Similarity 35.7%; Pred. No. 3.3e-37;
Matches 136; Conservative 77; Mismatches 131; Indels 37; Gaps 10;

```
Qy       5 NKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINS 64
           ||::||:| ::||||:|| ::: |   :   |::|   :|: |::||:: : :  |: |
Db       4 NKRVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTV--SNLYQESNPIVSEVTQYLYLKS 61

Qy      65 FKYFRNMYKGFYYSRPDKLDKCFYK-----YYGL-NKLINLLIKE-KPDLILLTFPTPVM 117
           |   :  |: |||     :|| : |     |: : || :  |: | :||:|: |||  |:
Db      62 FSIGKQFYRLFYYG----VDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVV 117

Qy     118 SVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGI 177
                :   ||   ||||: ||| |:     :|||||   |:  :::|  || ||:|||
Db     118 PEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKITGI 177

Qy     178 PIDNKFET-----PINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSANA 232
           ||  :||      || :|     || |:|: :|: ||| || |    :  : | |
Db     178 PIRPQFEESMPVGPIYKKY-----NLSPNKKVLLIMAGAHGVLKNVKELCEN-LVKDDQV 231
```

```
Qy    233 QVVMICGKSKELKRSLTA-------KFKLTRMYLILGYTKHMNEWMASSQLMITKPGGIT 285
          |||::|||:  || ||:|       | |      :||| : ::|    :  ||||||||| 
Db    232 QVVVVCGKNTALKESLSALEAENGDKLK------VLGYVERIDELFRITDCMITKPGGIT 285

Qy    286 ITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTNGNEQLTNM 345
          :||  |  :|:|   | |||| ||| :||::|   : :  || :: | ||    : |  |
Db    286 LTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDTLHRM 345

Qy    346 ISTMEQDKIKYATQTICRDLL 366
             ::   :  ::: |  |:|
Db    346 KKNIKDLHLANSSEVILEDIL 366


RESULT   2
C84499
probable monogalactosyldiacylglycerol synthase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84499
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: C84499
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-464 <STO>
A;Cross-references: GB:AE002093; NID:g4734000; PIDN:AAD28678.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g11810
A;Map position: 2


  Query Match             15.1%;  Score 309;  DB 2;  Length 464;
  Best Local Similarity   25.4%;  Pred. No. 9.7e-16;
  Matches  104;  Conservative   83;  Mismatches  178;  Indels   44;  Gaps   16;

Qy      6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSF 65
          | :||:    | ||    ::| :       |   :|  |:: |         | :|
Db     72 KTVLILMSDTGGGHRASAEAIRDAFKIEFGDDYRIIIKDVWKE--------YTGWPLNDM 123

Qy     66 -KYFRNMYK-------GFYYSRPDKLDKCFYK----YYGLNKLINLLIKEKPDLILLTFP 113
           : :: | |        |: : |  : | :      ||   ::   |:: |||:|:   |
Db    124 ERQYKFMVKHVGLWSVAFHGTSPKWIHKSYLSALAAYYA-KEIEAGLMEYKPDIIISVHP 182

Qy    114 ----TPVMSVLTEQFNINIPVATVMTDYR-LHKNWITPYSTRYYVATKETKQDFIDVGID 168
              |:  :  :  :   ||:||    |: |     :| |  :||  :  :  |:|
Db    183 LMQHIPLWVMKWQGLHKKVIFVTVITDLNTCHRTWFHHGVSRCYCPSKEVAKRALVDGLD 242

Qy    169 PSTVKVTGIPIDNKF-ETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMIT--DI 225
          | ::| |:|:   |  | :|: :   :  :| :   :|:  |  |:     | :   |
Db    243 DSQIRVFGLPVRPSFPRTILNKNELRKELEIDLNLPAVLLMGGGEGMGPVQKTALALGDS 302

Qy    226 L--AKSAN--AQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLMITK 280
          |  :| :|   |:::|||::| |  :| : ::|:     : |:   | :|| :   :|||
Db    303 LYNSKESNPIGQLIVICGRNKVLASTLASHEWKIP--VKVRGFETQMEKWMGACDCIITK 360

Qy    281 PGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVAS-LTNGN 339
          |  || |     :|:|  :  ||||  |  |  : | |    :|:|  ||||   :|
Db    361 AGPGTIAEALICGLPIILNDYIPGQEKGNVPYVVDNGAGVFTRSPKETAKIVADWFSNNK 420

Qy    340 EQLTNMISTMEQDKIKYATQTICRDLLDLIGHSSQPQEIYGKVPLYARF 388
          |:|      | :: :| :      |::  | | || |:  ::||:  |
Db    421 EEL----KKMSENALKLSQPEAVFDIVKDIHHLSQQQQ---RIPLFNEF 462


RESULT   3
F75439
probable cell wall synthesis protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C;Accession: F75439
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: F75439
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-411 <WHI>
A;Cross-references: GB:AE001958; GB:AE000513; NID:g6458805; PIDN:AAF10649.1; PID:g645
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR1076
A;Map position: 1


  Query Match             14.6%;  Score 299;  DB 2;  Length 411;
  Best Local Similarity   25.3%;  Pred. No. 4.7e-15;
  Matches   84;  Conservative   63;  Mismatches  153;  Indels   32;  Gaps    8;

Qy      7 KILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSFK 66
          : | :: | | || |  |::     :  :: |   |||          |     |    :
Db     29 RALFMSVSLGAGHDQAQQAVKQAFAERGVELLGA-EHDSVEYLSTFERSFTVDLYEFELR 87

Qy     67 YFRNMYKGFYY----SRPDKLDKCFYKYYGLNKLINLLIKEKPDLILLTF--PTPVMSVL 120
          |   :|:|||:     :|  :    : : |:     : | : :|:::: :|  |  |   |
Db     88 YAPWLYRGFYWLTDQDQPWNIISRMFTWLGMGAFKDELRELRPEVVINSFWAPAAVCDTL 147

Qy    121 TEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGIPID 180
            |         ::|||| | :|    :    ||::||::  :: |: |  |:||||||
Db    148 RAQTGQRFLNCLIVTDYRAHLHWARRETDLLMVASEETRRQMLERGVRPEQVEVTGIPIS 207

Qy    181 NKFETPINQKQWLI------DNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSAN--- 231
            |   :  :| :      : :| |    :|:| |     :|      |:| :  |
Db    208 PAFREVLAADRWALRAELFSEMSLRPGVPLLLLSGG----GRGHYAAAADVLTELGNLGR 263

Qy    232 -AQVVMIC---GKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITIT 287
           ||::     |:  |     |            ||: : :   :|:| |:: | ||:|:
Db    264 AVQVLVPASRQGEGTETIGGATVHH--------LGFRRDLPRLLAASDLVVGKAGGLTVA 315

Qy    288 EGFARCIPMIFLNPAPGQELENAFYFEEKGFG 319
          |  |  :|::   | ||||  || : |  | |
Db    316 EATALGVPLVIYAPIPGQEEHNADFLERHGAG 347


RESULT   4
T52269
1,2-diacylglycerol 3-beta-galactosyltransferase (EC 2.4.1.46) [imported] - Arabidopsi
N;Alternate names: monogalactosyldiacylglycerol synthase
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C;Accession: T52269
R;Awai, K.; Shimojima, M.; Masuda, T.; Takamiya, K.I.; Ohta, H.
submitted to the EMBL Data Library, July 1998
A;Description: cDNA cloning of an Arabidopsis monogalactosyldiacylglycerol synthase r
A;Reference number: Z26011
A;Accession: T52269
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-468 <AWA>
A;Cross-references: EMBL:AJ000331; PIDN:CAA04005.1
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: mgd
C;Keywords: glycosyltransferase; hexosyltransferase


  Query Match             14.5%;  Score 296.5;  DB 2;  Length 468;
  Best Local Similarity   25.4%;  Pred. No. 8.7e-15;
  Matches  100;  Conservative   78;  Mismatches  173;  Indels   43;  Gaps   14;
```

```
Qy       4 QNKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYIN 63
Db      67 RTKNVLILMSDTGGGHRASAEAIRDAFKIEFGDKYRVIVKDVWKE--------YTGWPLN 118

Qy      64 ----SFKYF---RNMYK-GFYYSRPDKLDKCFY----KYYGLNKLINLLIKEKPDLILLT 111
Db     119 DMERSYKFMVKHVQLWKVAFHSTSPKWIHSCYLAAIAAYYA-KEVEAGLMEYKPEIIISV 177

Qy     112 FP----TPVMSVLTEQFNINIPVATVMTDYR-LHKNWITPYSTRYYVATKETKQDFIDVG 166
Db     178 HPLMQHIPLWVLKWQELQKRVLFVTVITDLNTCHPTWFHPGVNRCYCPSQEVAKRALFDG 237

Qy     167 IDPSTVKVTGIPIDNKFETPINQKQWL-IDNNLDPDKQTILMSAGAFGV------SKGFD 219
Db     238 LDESQVRVFGLPVRPSFARAVLVKDDLRKELEMDQDLRAVLLMGGGEGMGPVKETAKALE 297

Qy     220 TMITDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLMI 278
Db     298 EFLYDKENRKPIGQMVVICGRNKKLASALEAIDWKIP--VKVRGFETQMEKWMGACDCII 355

Qy     279 TKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIV----AS 334
Db     356 TKAGPGTIAESLIRSLPIILNDYIPGQEKGNVPYVVENGAGVFTRSPKETARIVGEWFST 415

Qy     335 LTNGNEQLTNMISTMEQDKIKYATQTICRDLLDL 368
Db     416 KTDELEQTSDNARKLAQPE---AVFDIVKDIDEL 446
```

RESULT 5
T05092
probable 1,2-diacylglycerol 3-beta-galactosyltransferase (EC 2.4.1.46) - Arabidopsis tha
N;Alternate names: protein F28M20.30
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T05092
R;Bevan, M.; Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Hoheisel, J.; Mewes,
submitted to the Protein Sequence Database, November 1998
A;Reference number: Z15398
A;Accession: T05092
A;Molecule type: DNA
A;Residues: 1-533 <BEV>
A;Cross-references: EMBL:AL031004
A;Experimental source: cultivar Columbia; BAC clone F28M20
C;Genetics:
A;Map position: 4
A;Introns: 175/3; 233/2; 287/2; 334/3; 409/3; 430/3; 455/3
A;Note: F28M20.30
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 13.9%; Score 284; DB 2; Length 533;
Best Local Similarity 25.5%; Pred. No. 9.2e-14;
Matches 103; Conservative 66; Mismatches 185; Indels 50; Gaps 13;

```
Qy       6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSF 65
Db     142 KKVLILMSDTGGGHRASAEAIRAAFNQEFGDEYQVFITDLWTDHTP--------WPFNQL 193

Qy      66 KYFRN-------MYKGFYYSRPDKL----DKCFYKYYGLNKLINLLIKEKPDLILLTFP- 113
Db     194 PRSYNFLVKHGTLWKMTYYGTSPRIVHQSNFAATSTFIAREIAQGLMKYQPDIIISVHPL 253

Qy     114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
Db     254 MQHVPLRVLRSKGLLKKIVFTTVITDLSTCHPTWFHKLVTRCYCPSTEVAKRAQKAGLET 313

Qy     170 STVKVTGIPIDNKFETPINQKQWL-----IDNNLDPDKQTILMSAGAFGV------SKGF 218
Db     314 SQIKVYGLPVRPSFVKPVRPKVELRRELGMDENL----PAVLLMGGGEGMGPIEATARAL 369

Qy     219 DTMITDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLM 277
```

```
Db     370 ADALYDKNLGEAVGQVLIICGRNKKLQSKLSSLDWKIP--VQVKGFITKMEECMGACDCI 427

Qy     278 ITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTN 337
Db     428 ITKAGPGTIAEAMIRGLPIILNGYIAGQEAGNVPYVVENGCGKFSKSPKEISKIVADWFG 487

Qy     338 GNEQLTNMISTMEQDKIKYATQ----TICRDLLDLI-GHSSQPQ 376
Db     488 ---PASKELEIMSQNALRLAKPEAVFKIVHDMHELVRKKNSLPQ 528
```

RESULT 6
T10478
probable 1,2-diacylglycerol 3-beta-galactosyltransferase (EC 2.4.1.46) precursor, chl
N;Alternate names: monogalactosyldiacylglycerol synthase
C;Species: Cucumis sativus (cucumber)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C;Accession: T10478
R;Shimojima, M.; Ohta, H.; Iwamatsu, A.; Masuda, T.; Shioi, Y.; Takamiya, K.
Proc. Natl. Acad. Sci. U.S.A. 94, 333-337, 1997
A;Title: Cloning of the gene for monogalactosyldiacylglycerol synthase and its evolut
A;Reference number: Z17042; MUID:97144442
A;Accession: T10478
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-525 <SHI>
A;Cross-references: EMBL:U62622; NID:g1805253; PIDN:AAC49624.1; PID:g1805254
A;Experimental source: cv. Aonagajibai; 5 day old seedlings
C;Genetics:
A;Genome: nuclear
C;Function:
A;Description: catalyzes the formation of monogalactosyldiacylglycerol, a major struc
A;Pathway: membrane lipid biosynthesis
C;Keywords: chloroplast; glycosyltransferase; hexosyltransferase
F;1-103/Domain: transit peptide (chloroplast) #status predicted <TNP>
F;104-525/Product: 1,2-diacylglycerol 3-beta-galactosyltransferase #status predicted

Query Match 13.8%; Score 283; DB 2; Length 525;
Best Local Similarity 25.3%; Pred. No. 1.1e-13;
Matches 105; Conservative 65; Mismatches 185; Indels 60; Gaps 13;

```
Qy       6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSF 65
Db     136 KRVLILMSDTGGGHRASAEAIKAAFNEEFGNNYQVFITDLWTDHTP--------WPFNQL 187

Qy      66 KYFRN-------MYKGFYYSRPDKL----DKCFYKYYGLNKLINLLIKEKPDLILLTFP- 113
Db     188 PRSYNFLVKHGTLWKMTYYVTAPKVIHQSNFAATSTFIAREVAKGLMKYRPDIIISVHPL 247

Qy     114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
Db     248 MQHVPIRILRSKGLLNKIVFTTVVTDLSTCHPTWFHKLVTRCYCPSTEVAKRALTAGLQP 307

Qy     170 STVKVTGIPIDNKFETPIN-----QKQWLIDNNLDPDKQTILMSAGAFGV------SKGF 218
Db     308 SKLKVFGLPVRPSFVKPIRPKIELRKELGMDENL----PAVLLMGGGEGMGPIEATAKAL 363

Qy     219 DTMITDILAKSANAQVVMICGKSKELK---RSLTAKFKLTRMYLILGYTKHMNEWMASSQ 275
Db     364 SKALYDENHGEPIGQVLVICGHNKKLAGRLRSIDWKVPVQ----VKGFVTKMEECMGACD 419

Qy     276 LMITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASL 335
Db     420 CIITKAGPGTIAEAMIRGLPIILNDYIAGQEAGNVPYVVENGCGKFSKSPKEIANIVAKW 479

Qy     336 TNGNEQLTNMISTMEQDKIKYATQ----TICRDLLDLIGHSSQPQEIYGKVPLYA 386
Db     480 FGPK---ADELLIMSQNALRLARPDAVFKIVHDLHELVKQRS-------FVPQYS 524
```

RESULT 7
B69860

```
conserved hypothetical protein ykoN - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: B69860
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033
A;Accession: B69860
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-373 <KUN>
A;Cross-references: GB:Z99110; GB:Z99111; GB:AL009126; NID:g2633699; PIDN:CAB13208.1; PI
A;Experimental source: strain 168
C;Genetics:
A;Gene: ykoN

  Query Match             12.0%;  Score 245;  DB 2;  Length 373;
  Best Local Similarity   23.5%;  Pred. No. 4.9e-11;
  Matches   88;  Conservative   75;  Mismatches  168;  Indels   44;  Gaps   14;

Qy       6 KKILIITG-SFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINS 64
           | |||    |   ||  |  ::  :|       |:  : |:|  ::  |  :    |:
Db       2 KNILIFPFLSISTGHHHVADALQAELESQG---LAAEKIDIFSHSYRRLEKLSSVAYLKW 58

Qy      65 FKYFRNMYKGFYYSRPDKL--------DKCFYKYYGL--NKLINLLIKEKPDLILLT--F 112
           :||   | | |     :|         || :: |  :   :: ::| :::||:   |
Db      59 IQYFPKTYSGIY-----RLLACGEFQHDKRYFMYECVFTQQMRHILQEKQPDIAFCTHAL 113

Qy     113 PTPVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTV 172
           |: ::: |  ::  |: |  | ||: ::: |       ::| : | |:  |  ||| : :
Db     114 PSYLLNRLKPEYP-NLTVVNVYTDFFVNQLWGRKNIDYHFVPSTEVKKQLISEGIDQNNI 172

Qy     173 KVTGIPIDNKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSANA 232
           :|||||:   ||          |         ||::: |: ||  |    : : |:
Db     173 YLTGIPVHQNFEME------SADTLQHHPPYTIIITGGSMGVG-GILKWVQE-LSPGGKI 224

Qy     233 QVVMICGKSKELKRSL-TAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITITEGFA 291
             ::||::::|   : :    |       |     ||     :  ::|||||:||:|
Db     225 LYKILCGRNEKLYSYVKSLHHPLIEAIPYLHSKAEMNRLYEQATGIMTKPGGVTISECLQ 284

Qy     292 RCIPMIFLNPAPGQELENAFYFEEKGFGKIAD----TPEEAIKIVASLTNGNEQL----- 342
           : :|:   :  ||||  |     |:    : |      ::| : :|:    |||:
Db     285 KRLPVFIYHALPGQEEMNLNLLHERKL--VTDMRNWDMQKAEEYIAAFFQSNEQMKEYKK 342

Qy     343 --TNMISTMEQDKIK 355
               :  |   |||
Db     343 HVNGYLGEMSDRKIK 357


RESULT   8
E83894
hypothetical protein BH1957 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C;Accession: E83894
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20263314
A;Accession: E83894
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-374 <STO>
A;Cross-references: GB:AP001513; GB:BA000004; NID:g10174345; PIDN:BAB05676.1; GSPDB:G
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH1957

  Query Match             11.2%;  Score 230;  DB 2;  Length 374;
  Best Local Similarity   21.8%;  Pred. No. 6.7e-10;
  Matches   83;  Conservative   82;  Mismatches  165;  Indels   50;  Gaps   14;

Qy       9 LIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSFKYF 68
           || : | |||| |  :::  |:   |  :   |  | |    | |      |:|  |
Db       6 LIFSASIGNGHNQAAKAL--QVEFQNKGYQPEI-IDTFYSLSPALHKFMLTSYVNLLKVG 62

Qy      69 RNMYKGFYYSRPDK------LDKCFYKYYGLNKLINLLIKEKPDLILLTFP--TPVMSVL 120
             :::  |: : :|      ||:  :  : :  |   :   :   :: | |  |  :  |
Db      63 PRIWQKIYF-QAEKYPLFLFLDQ--FATFFVESLHATVKSNRCSFLVSTHPFVTAFLVRL 119

Qy     121 TEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDV-GIDPSTVKVTGIPI 179
             :  :|:|: ||:||: ||  :: |    |: :      ||  :  :       |||||
Db     120 KSKKQLNLPLYTVITDFVLHPAYLRPEIDGYFTSDPNF-TDFAKLNNVSDDRFFPTGIPI 178

Qy     180 DNKFETPINQKQWLIDNN--LDPDKQTILMSAGAFGVSKGFDTMITDILAKSANAQVVMI 237
           |     |:| :| : |:  ||| :: :::: |  |::  :  :|  :       |:: :
Db     179 PNL--ESIDQPKWKVRNDLGLDPQRKVLIIAGGGIGLT-NYAQVIRALECLPEPIQLLCM 235

Qy     238 CGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITITEGFARCIPMI 297
           | : ::|  : :: |       :: :|     :: :|  :::| ||:|: |      |:|
Db     236 IGHNYQVKEKI-SRIKSKHELKVIEFTDKFLLYLKASDAILSKAGGLTMAESLVCETPII 294

Qy     298 FLNPAPGQELENAFY----------------------FEEKGFGKIADTPEEAIKIVA 333
           | || |  || :                          :||  || : :   :  |  |
Db     295 IHQPVPGHEEHNAKFLIDAGAALRVKGSKEIPTTIKRVLYEEACFGPMIENARKLKKPNA 354

Qy     334 SLTNGNEQLTNMISTMEQDK 353
           :    || :  |:  ::: : '
Db     355 A----NEIVEQMLLLVKEQQ 370


RESULT   9
F70195
UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 24-Nov-1999
C;Accession: F70195
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943
A;Accession: F70195
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-363 <KLE>
A;Cross-references: GB:AE001176; GB:AE000783; NID:g2688699; PIDN:AAC67113.1; PID:g268
A;Experimental source: strain B31
C;Superfamily: murG protein

  Query Match              6.6%;  Score 136;  DB 2;  Length 363;
  Best Local Similarity   21.2%;  Pred. No. 0.0082;
  Matches   86;  Conservative   68;  Mismatches  155;  Indels   96;  Gaps   18;

Qy       5 NKKILIITGSFGNGHMQVTQSIVNQLNDMN-------LDHLSVIEHDLFMEAHPI--LTS 55
           ||||:  ||    ||:    ||: :| : :        :   : ||  |  |   |  ::
Db       3 NKKIIFFTGGGTGGHVFPGISIIQKLKEFDNEIEFFWIGKKNSIEEKLIKEQDNIKFISI 62
```

```
Qy      56 ICKKW--YINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDLILLT-- 111
           | |    |  ||| | : :|                 |: |   :| | || ||  |
Db      63 PCGKLRRYF-SFKNFTDFFK--------------VILGIIKSFYVLKKYKPQLIYATGG 106

Qy     112 -FPTPV--------MSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDF 162
            ||         :  :| : :::  :|| :          : ::   ::: ||::: |
Db     107 FVSTPAIIASSLLKIKSITHEMDLDPGLATKIN---------SKFANNIHISFKESEKYF 157

Qy     163 IDVGIDPSTVKVTGIPIDNKFETPINQKQWLIDNNLDPDKQTILMSA--------GAFGV 214
           :       :  || ||  :|             |:|| : |                |
Db     158 KNY----KNIIYTGSPIRREF-------------LNPDPKIIKQLTQNTNKPIISILGG 199

Qy     215 SKGFDTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASS 274
           | | : :    |    :|::  |    | |   |: |  | | :      : |   :  |
Db     200 SLGANALNNLALCIKKDAEIYFIHQSGKNL-NDLSEKNYLRRQFF---NAEEMASIVKFS 255

Qy     275 QLMITKPGGITITEGFAR---CIPMI-FLNPAPGQELENA---------FYFEEKGFGKI 321
           |:|:: |   | | ||    |  :| |   : | :::||          | :|     |
Db     256 NLIISRAGAGAIKE-FANAGACAILIPFKKGSRGDQIKNAKLLTNQNACIYIDEDEILNI 314

Qy     322 ADTPEEAIKIVASLTNGNEQLTNMISTMEQDKIKYATQTICRDLL 366
              :||:       |:: ::   :::   |::: | : |:
Db     315 -----NILKIIKKTLKDREKINSLKENIKKFNNKHSSTLIAKLLI 354


RESULT  10
S38160
NUP133 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YKR082w; protein YKR402
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000
C;Accession: S38160; S42011; S51915; S39123
R;Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo,
submitted to the Protein Sequence Database, March 1994
A;Reference number: S38158
A;Accession: S38160
A;Molecule type: DNA
A;Residues: 1-1157 <BAL>
A;Cross-references: EMBL:Z28307; NID:g486562; PID:g486563; MIPS:YKR082w
A;Experimental source: strain S288C
R;Garcia-Cantalejo, J.; Baladron, V.; Esteban, P.F.; Santos, M.A.; Bou, G.; Remacha, M.A
Yeast 10, 231-245, 1994
A;Title: The complete sequence of an 18,002 bp segment of Saccharomyces cerevisiae chrom
A;Reference number: S42009; MUID:94262327
A;Accession: S42011
A;Molecule type: DNA
A;Residues: 1-1157 <GAR>
A;Cross-references: EMBL:Z27116; NID:g415899; PIDN:CAA81633.1; PID:g415902
A;Experimental source: strain S288C
R;Doye, V.; Wepf, R.; Hurt, E.C.
EMBO J. 13, 6062-6075, 1994
A;Title: A novel nuclear pore protein Nup133p with distinct roles in poly(A)(+) RNA tran
A;Reference number: S51915; MUID:95112817
A;Accession: S51915
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1157 <DOY>
A;Cross-references: EMBL:X80066; NID:g624033; PIDN:CAA56372.1; PID:g624034
C;Genetics:
A;Gene: SGD:NUP133
A;Cross-references: SGD:S0001790; MIPS:YKR082w
A;Map position: 11R
C;Keywords: nucleus; transmembrane protein
F;413-429/Domain: transmembrane #status predicted <TM1>
F;641-657/Domain: transmembrane #status predicted <TM2>


  Query Match             6.2%;  Score 127.5;  DB 2;  Length 1157;
  Best Local Similarity   20.9%;  Pred. No. 0.19;
  Matches  104;  Conservative   67;  Mismatches  147;  Indels  179;  Gaps   27;

Qy       2 VTQNKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIE--HDLF---------MEAH 50
             :| ||      | |    :  | |   :| |:|: : :::|   ||:          ::|

Db     285 ITTNK------GIFQTWQLSATNSHPTKLIDVNI-YEAILESLQDLYPFAHGTLKIWDSH 337

Qy      51 PI--------LTSI----CKK-WYINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLI 97
           |:         |:||    | : :|| |   | :    |             :  | ||  :
Db     338 PLQDESSQLFLSSIYDSSCNETYYILSTIIFDSSSNSF----------TIFSTYRLNTFM 387

Qy      98 NLL--IKEKPDL---------------ILLTFPTPVMSVLTEQFNINIPVATVMTDYRL 139
             :   | || :                ||: ||  |  |:|:   :|   : : : | :
Db     388 ESITDTKFKPKIFIPQMENANDTNEVTSILVMFPNAV--VITQ---VN---SKLDSSYSM 439

Qy     140 HKNWITPYSTR--------------YYVATKE--TKQDFI----------DVGIDPSTV- 172
             : |    | |                || ||:   | |:           :||   | |
Db     440 RRKWEDIVSLRNDIDIIGSGYDSKSLYVLTKQMGVLQFFVKENEETNSKPEVGFVKSHVD 499

Qy     173 ------KVTGIPIDNKFETPINQKQWLIDNNLDPDKQTILMSAGAF-----------GV 214
                 |:   |||       |:  |  |:::|    : |  | | :             |
Db     500 QAVYFSKINANPIDFNLPPEISLDQESIEHDLKLTSEEIFHSNGKYIPPMLNTLGQHLSV 559

Qy     215 SKGFDTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLTRMYL--------------- 259
             | |       :||: | ::       | |||  |  ||::    :
Db     560 RKEFFQNFLTFVAKNFNYKI------SPELKLDLIEKFEILNCCIKFNSIIRQSDVLNDI 613

Qy     260 ----ILGYTKHMNEWMASSQLMITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEE 315
               :  |    || : :  ::|  |              |:||      ::  |   |
Db     614 WEKTLSNYNLTQNEHLTTKTVVINSPD----------VFPVIF------KQFLNHVVF-- 655

Qy     316 KGFGKIADTPEEAIKIVASLTNGNEQL--TNMISTMEQDKI-KYATQTICRDLLDLIGHS 372
                       ::    | | :|  ||:|:    | | :   :||  :||:|
Db     656 --------------VLFPSQNQNFKLNVTNLINLCFYDGILEEGEKTIRYELLEL---- 696

Qy     373 SQPQEI-YGKVPLYARF 388
             | |:   |:| :  |
Db     697 -DPMEVDTSKLPWFINF 712


RESULT  11
E83970
UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)pyrophos murG [imported] - Bac
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 08-Dec-2000
C;Accession: E83970
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A;Reference number: A83650; MUID:20263314
A;Accession: E83970
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-363 <STO>
A;Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BAB06284.1; GSPDB:G
A;Experimental source: strain C-125
C;Genetics:
A;Gene: murG
C;Superfamily: murG protein


  Query Match             6.1%;  Score 124;  DB 2;  Length 363;
  Best Local Similarity   19.8%;  Pred. No. 0.066;
  Matches   77;  Conservative   82;  Mismatches  149;  Indels   80;  Gaps   18;

Qy       9 LIITGSFGNGHMQVTQSIVNQL--NDMNLDHLSV-----IEHDL-------FMEAHPILT 54
             ::::|    ||:    : :|::   |  || | :     :| ::       |   |  :|
Db       3 IVVSGGGTGGHIYPALAFINEMKKRDERLDVLYIGTERGLESEIVPREGIPFQTIH--IT 60

Qy      55 SICKKWYINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKE-KPDLILLTFP 113
             :|  : : |       :|                    |    |: | |||::: |
Db      61 GFQRKLSMENVKTVVRFLRG------------------TKRAKALLNEFKPDVVIGTGG 101

Qy     114 TPVMSVLTEQFNINIPVATVMTDYR----LHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
                 |:     : ||  ||: :       |   ::: |  |  :  || :  |
Db     102 YVCGPVVYAAAKLKIP--TVIHEQNSVPGLTNKFLSRYVDRIAICFKEAEAFF-----PK 154

Qy     170 STVKVTGIPIDNKFETPINQKQWLIDNNLDPDKQTILM---SAGAFGVSKGFDTMITDIL 226
```

```
          : |  || |  ::  :  |::: |    : |:|:|:|:   | ||  ::  | ::::|:
Db    155 NKVVFTGNPRASEVMSG-NREEGLRSLGIKPNKKTVLIVGGSRGARPINDAFMSILSDVK 213

Qy    227 AKSANAQVVMICG-----KSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKP 281
          ||    | | : |     : :|  :|:     :    ::  :  :| : :::  |:: :
Db    214 AKP--YQFVYVTGTVHYERVQEQMKSIGQPENV----IVQPFIHNMPDVLSAVDLIVARA 267

Qy    282 GGITITEGFARCIPMIFLNPAP----GQELENAFYFEEKGFGKIADTPEEAIKIVASLTN 337
          |  |: |  |  :| | : |:|      : :||    :|         : ||    |
Db    268 GATTLAEITALGLPSILI-PSPYVTNNHQEKNAAALSKK---------DAAILRKESELT 317

Qy    338 GNEQLTN----MISTMEQDKIKYATQTI 361
          |:  | :    |::    | :| | : :
Db    318 GDRLLEDIDDIMVTPGRLDAMKQAAKAL 345


RESULT 12
C70401
phospho-N-acetylmuramoyl-pentapeptide transferase - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C;Accession: C70401
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666
A;Accession: C70401
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-344 <AQF>
A;Cross-references: GB:AE000727; NID:g2983623; PIDN:AAC07193.1; PID:g2983625; GB:AE00065
A;Experimental source: strain VF5
C;Genetics:
A;Gene: murG
C;Superfamily: murG protein


  Query Match             6.0%;  Score 123.5;  DB 2;  Length 344;
  Best Local Similarity   20.7%;  Pred. No. 0.067;
  Matches   80;  Conservative   72;  Mismatches  131;  Indels  103;  Gaps   19;

Qy      2 VTQNKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWY 61
          | : :|:  |    |         |  :| |:      :    ||::  |:
Db     23 VLKKEKVKFIGSKRG---------IEYELKDL------IKTEKLFLDVEPLRE------- 60

Qy     62 INSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDLILL----TFPTPVM 117
                 || |:        || |  :|:    : ||  :||    ::     : |  :
Db     61 -------RNFYQ--------KL-KAIWKFLKAQEEINEFLKEDYRALIFGGYASLPLGIN 104

Qy    118 SVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKE---------TKQDFIDVGID 168
          :||  :          : :|:    |  |   :: |           ||: |
Db    105 TVLRRK------------ELFIHEQNSIPSKTNKILSKKAKKVLITFNYTKRFF------ 146

Qy    169 PSTVKVTGIPIDNKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAK 228
          |  |:| |:||  : :  : :|:      |:||| |:|:    || |:|   :  : ||:
Db    147 PEGVRV-GLPIRKELKKKLPKKEVKKRFGLEPDKITVLI----FGGSQG--ALFLNELAR 199

Qy    229 SANA------QVVMICGK--SKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITK 280
             :      ||::: ||   :: |     ||:      :: ::  |    ::| : |::
Db    200 DLKSVLPKEFQVILLTGKIHYEKFKNLEGEKFR------VMPFSLDMGLIYSASDVAISR 253

Qy    281 PGGITITEGFARCIPMIFLNPAPGQELENAFY----FEEKGFGKIADTPE-------EAI 329
          |  || |     :| :|: | |    :: ||     |: | | :  |         |:
Db    254 AGAGTINELSHFGVPSVFV-PYPYAVDDHQFYNAKEIEKLGGGLVLRQEEAKPDKVLSAL 312

Qy    330 K-IVASLTNGNEQLTNMISTMEQDKI 354
          | || :|   :| :    :  ::::
Db    313 KEIVKNLERYSENIKKFFAEGAEERM 338


RESULT 13
S35481
SRPM54 protein - Mycoplasma mycoides
C;Species: Mycoplasma mycoides
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-Dec-1999
C;Accession: S35481; S27591
R;Samuelsson, T.
Nucleic Acids Res. 20, 5763-5770, 1992
A;Title: A Mycoplasma protein homologous to mammalian SRP54 recognizes a highly conse
A;Reference number: S35480; MUID:93087189
A;Accession: S35481
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-447 <SAM>
A;Cross-references: EMBL:M91593; NID:g150208; PIDN:AAA25441.1; PID:g150210
C;Genetics:
A;Genetic code: SGC3
C;Superfamily: signal recognition particle 54K protein


  Query Match             6.0%;  Score 122;  DB 2;  Length 447;
  Best Local Similarity   22.9%;  Pred. No. 0.13;
  Matches   80;  Conservative   52;  Mismatches  143;  Indels   74;  Gaps   16;

Qy     13 GSFGNGHMQ--VTQSIVNQ-LNDMNLD------HLSVIEHDLFMEAHPILTSICKKWYIN 63
          | | :  ||  : ::: |  ||: |:        ||::| |: :||          |  ||
Db      4 GDFLSKRMQKSIEKNMKNSTLNEENIKETLKEIRLSLLEADVNIEA--------AKEIIN 55

Qy     64 SFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDLILLTFPTPVMSVLTEQ 123
          : |       | | |       :   |      :|:|:| ||    | :   |: || |  :
Db     56 NVK---QKALGGYISEGASAHQQMIKIVH-EELVNILGKENAPLDINKKPSVVMMVGLQ- 110

Qy    124 FNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGID---PSTVKVTGIPID 180
                               | :   |:  |: |:  : ||:|    |  ::   : :
Db    111 ----------------GSGKTTTANKLAYLLNKKNKKKVLLVGLDIYRPGAIEQL-VQLG 153

Qy    181 NKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSANAQVVMICGK 240
          |  | : :|        || |       |  :      || :| |   :    ||:|
Db    154 QKTNTQVFEK-----GKQDPVKTA--EQALEYAKENNFDVVILDTAGRLQVDQVLM---- 202

Qy    241 SKELKRSLTAKFKLTRMYL---------ILGYTKHMNEWMASSQLMITK-----PGGITI 286
           |||  :|  |     : |         |:  |   |: :  | :::||      || |:
Db    203 -KELD-NLKKKTSPNEILLVVDGMSGQEIINVTNEFNDKLKLSGVVVTKLDGDARGGATL 260

Qy    287 TEGFARCIPMIFLNPAPGQELENAFYFEE-----KGFGKIADTPEEAIK 330
          :  :   :|: |:    |     ||| :       | | |    | |::
Db    261 SISYLTKLPIKFIGEGEGYNALAAFYPKRMADRLMGMGDIETLFERAVE 309


RESULT 14
T44332
hypothetical protein wblE [imported] - Vibrio cholerae
C;Species: Vibrio cholerae
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T44332
R;Yamasaki, S.; Shimizu, T.; Hoshino, K.; Ho, S.T.; Shimada, T.; Nair, G.B.; Takeda,
Gene 237, 321-332, 1999
A;Title: The genes responsible for O-antigen synthesis of Vibrio cholerae O139 are cl
A;Reference number: Z22749; MUID:99453293
A;Accession: T44332
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-359 <YAM>
A;Cross-references: EMBL:AB012957; NID:g4115688; PIDN:BAA33636.1; PID:g3721686
A;Experimental source: strain O22
C;Genetics:
A;Note: wblE
C;Superfamily: probable hexosyltransferase ytxN


  Query Match             5.9%;  Score 121;  DB 2;  Length 359;
  Best Local Similarity   19.3%;  Pred. No. 0.11;
  Matches   79;  Conservative   76;  Mismatches  139;  Indels  116;  Gaps   19;

Qy     10 IITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSFKYFR 69
```

```
        :||   | |     :::|: : | :          ||:: : |              ||
Db      11 VITSDSGGG----AENLVSNMLDSD---------GLFLQ-YAI--------------YFS 42

Qy      70 NMYKGFYYSRPDKLDKCFYKYYG-----------LNKLINLLIKEKPDLILLTFPT-PVM 117
           :   |  | ||: :      |:|           | | | :||::  :||:    | |:
Db      43 SFISG--YRRPNTI------YFGSNPRSLLNIFKLRKTIKMLIEKHGELIVHAHLTWPLF 94

Qy     118 SVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFID------------- 164
            |    ||:|: :              | :||      |  |  |:
Db      95 YVALGLFNLNVKLV-----------YTEHST----FNKRRKYPFLKFFERLVYSRYSRI 138

Qy     165 VGIDPSTVKVTGIPIDNKFETPINQKQWLIDN----------NLDPDKQTILMSAGAFG 213
           |||  |  |   : : : | : |      |:|          ||:  |:  |:| |:
Db     139 VGISNSVSKSIEVWLGSSFLSRI----VTINNGARFYSFKCRHNLELSKKIKLLSVGSLK 194

Qy     214 VSKGFDTMITDI--LAKSANAQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWM 271
             |||| |  :  : :  :   ::  | | :  : ::    :  :   :|::  : ::
Db     195 PLKGFDKTIKSLSKITQYIDKYTIVGTGPSYDTLKEISLDNHVDHLVDFVGWSNELEKFY 254

Qy     272 ASSQLMIT----KPGGITITEGFARCIPMIF-----LNPAPGQELENAF----YFEEKGF 318
           : :::     :  |:: || :  :|:|      ||    : | ::     : :|  :
Db     255 HEADILLIPSHWEGFGLSAVEGMSTGLPVIASNVDGLNEVVSKTLNSSILVENFLDENSW 314

Qy     319 GKIADTPEEAIKIVASLTNGNEQLTNMISTMEQDKIKYATQTICRDLLDL 368
             ||   |  | ::|        ||    :|   |::   : ::  ||
Db     315 ---ADAILEMKKKLSS------DYFNMSIESKQQAYKFSFDKMFKEYTDL 355
```

RESULT 15
C64664
transferase, peptidoglycan synthesis - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 18-Feb-2000
C;Accession: C64664
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467
A;Accession: C64664
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-353 <TOM>
A;Cross-references: GB:AE000621; GB:AE000511; NID:g2314301; PIDN:AAD08196.1; PID:g231430
C;Superfamily: murG protein

Query Match 5.7%; Score 117.5; DB 2; Length 353;
Best Local Similarity 20.6%; Pred. No. 0.2;
Matches 81; Conservative 59; Mismatches 169; Indels 85; Gaps 20;

```
Qy      11 ITGSFGNGHMQVTQSIVNQLNDMNLDHL---SVIEHDLFMEAHPILTSICKKWYINSFKY 67
           :||    ||: : :::  :|    :: :   |    |             |:|: ||
Db       5 LTGGGTGGHLSIAKALAIELEKQGIEAIYLGSTYGQD------------KEWFENS-PL 50

Qy      68 FRNMYKGFYYSRPDKLDKCFYKYYG---LNKLINLLIKE--KPDLILLTF--------PT 114
           |   |   :::    ::| |:|  |   |          ||  |   |  |        |
Db      51 FSERY---FFNTQGVVNKSFFKKIGSLFLQAKAAFKAKEILKKHQITHTISVGGFSAGPA 107

Qy     115 PVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGI-DPSTVK 173
           |:| :     ||:       |   :| |     ||   : : |  |       |
Db     108 SFASLLNK-----IPL------YIHEQNAIKGSLNRY--LSPKAKAVFSSYAFKDKGNHV 154

Qy     174 VTGIPIDNKF------ETPINQKQWLIDNNLDPDKQTILMSAGAFGVS--KGFDTMITDI 225
           :|  |: | |       | |                  ||   |: |       |  :
Db     155 LTSYPVQNAFFDFARTRTEIKH--------------ILFLGGSQGAKAINEFALLNAPK 199

Qy     226 LAKSANAQVVMICG-KSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGI 284
           | |    ::  |||  | |  |    :  |     :  :  :: | |  : | ::: |
Db     200 LTKQ-GIKITHICGPNSYEQVRFFYQELGLLDKIELFAFHNNITEIMHRADLCVSRAGAS 258

Qy     285 TITEGFARCIPMIFLNPAP----GQELENAFYFEEKGFGKIAD----TPEEAIKIVASLT 336
           :: |  |  :| ||: | |      :  |   ||::    :          |::  :::  |
Db     259 SVWELCANGLPTIFI-PYPFASNNHQYYNVLEFEKENLCYVVPQNELLPKKLFEVIRKLN 317

Qy     337 ----NGNEQLTNMISTMEQDKIKYATQTICRDLL 366
                ||: || : : ::|    |    :||   :|
Db     318 QKDDQGNKNLTTISNQLQQKIAKDGAKTIIETIL 351
```

Search completed: June 29, 2001, 09:00:20
Job time: 182 sec

```
OM protein - protein search, using sw model

Run on:          June 29, 2001, 09:04:10 ; Search time 16.33 Seconds
                                           (without alignments)
                                           820.202 Million cell updates/sec

Title:           US-09-668-788-4
Perfect score:   2048
Sequence:        1 MVTQNKKILIITGSFGNGHM..........SSQPQEIYGKVPLYARFFVK 391

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters:      93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_39:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	591.5	28.9	382	1	YPFP_BACSU	P54166 bacillus su
2	136	6.6	363	1	MURG_BORBU	O51708 borrelia bu
3	135	6.6	363	1	MURG_ENTFA	O07109 enterococcu
4	127.5	6.2	1157	1	N133_YEAST	P36161 saccharomyc
5	124	6.1	352	1	MURG_STRPN	Q9zha9 streptococc
6	123.5	6.0	344	1	MURG_AQUAE	O67238 aquifex aeo
7	122	6.0	447	1	SR54_MYCMY	Q01442 mycoplasma
8	119	5.8	531	1	UD17_RAT	Q64633 rattus norv
9	117.5	5.7	353	1	MURG_HELPY	O25770 helicobacte
10	116.5	5.7	353	1	MURG_HELPJ	Q9zk59 helicobacte
11	113	5.5	950	1	Y511_RICPR	Q9zd36 rickettsia
12	111.5	5.4	360	1	MURG_ENTHR	O07670 enterococcu
13	111.5	5.4	363	1	MURG_BACSU	P37585 bacillus su
14	111.5	5.4	534	1	UD14_HUMAN	P22310 homo sapien
15	110.5	5.4	531	1	UD15_RAT	Q64638 rattus norv
16	110.5	5.4	534	1	UD13_HUMAN	P35503 homo sapien
17	109.5	5.3	533	1	UD12_RAT	P20720 rattus norv
18	109	5.3	354	1	MURG_BUCAI	P57311 buchnera ap
19	108.5	5.3	535	1	UD11_MOUSE	Q63886 mus musculu
20	106.5	5.2	530	1	UD18_RAT	Q64634 rattus norv
21	106.5	5.2	531	1	UD13_RAT	Q64637 rattus norv
22	106	5.2	339	1	MURG_THEMA	Q9wy74 thermotoga
23	106	5.2	530	1	UDBF_HUMAN	P54855 homo sapien
24	106	5.2	530	1	UDBK_MACFA	O77649 macaca fasc
25	105.5	5.2	273	1	SP0A_CLOPA	P52940 clostridium
26	104.5	5.1	529	1	UD16_RAT	P08430 rattus norv
27	104	5.1	1196	1	BXCN_CLOBO	P46081 clostridium
28	103.5	5.1	484	1	YA62_METJA	Q58462 methanococc
29	102.5	5.0	532	1	UD14_RABIT	Q28612 oryctolagus
30	102.5	5.0	934	1	MSH2_HUMAN	P43246 homo sapien
31	102	5.0	530	1	UDBH_HUMAN	O75795 homo sapien
32	102	5.0	798	1	V120_HSV6G	P30003 human herpe
33	101.5	5.0	376	1	WECB_ECOLI	P27828 escherichia
34	101.5	5.0	535	1	UD11_RAT	Q64550 rattus norv
35	101.5	5.0	1248	1	TOPG_SULAC	Q08582 sulfolobus
36	101	4.9	1082	1	V120_HSV6U	P52437 human herpe
37	100.5	4.9	533	1	UD12_MOUSE	P70691 mus musculu
38	100.5	4.9	1033	1	T1R1_ECOLI	P10486 escherichia
39	100	4.9	511	1	CHLB_PINST	Q95666 pinus strob
40	100	4.9	531	1	UD16_MOUSE	Q64435 mus musculu
41	100	4.9	534	1	UD15_HUMAN	P35504 homo sapien
42	99.5	4.9	533	1	UD11_HUMAN	P22309 homo sapien
43	99	4.8	686	1	RECG_BORBU	O51528 borrelia bu
44	99	4.8	1004	1	SLPO_BACBR	P09333 bacillus br
45	98.5	4.8	274	1	RT02_PROWI	P46741 prototheca

ALIGNMENTS

```
RESULT   1
YPFP_BACSU
ID   YPFP_BACSU     STANDARD;      PRT;   382 AA.
AC   P54166;
DT   01-OCT-1995 (Rel. 34, Created)
DT   01-OCT-1995 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL 43.6 KDA PROTEIN IN CPSD-METB INTERGENIC REGION.
GN   YPFP.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96349105; PubMed=8760912;
RA   Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,
RA   Serror P.;
RT   "Sequence analysis of the Bacillus subtilis chromosome region between
RT   the serA and kdg loci cloned in a yeast artificial chromosome.";
RL   Microbiology 142:2005-2016(1996).
CC   -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L77246; AAA96624.1; -.
DR   EMBL; Z99115; CAB14110.1; -.
DR   SubtiList; BG11611; ypfP.
DR   InterPro; IPR001296; -.
DR   Pfam; PF00534; Glycos_transf_1; 1.
KW   Hypothetical protein; Transferase; Glycosyltransferase.
SQ   SEQUENCE   382 AA;  43561 MW;  8A8B98593DAC6041 CRC64;

  Query Match             28.9%;  Score 591.5;  DB 1;  Length 382;
  Best Local Similarity   35.7%;  Pred. No. 6.2e-38;
  Matches  136;  Conservative   77;  Mismatches  131;  Indels   37;  Gaps   10;

Qy          5 NKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINS 64
              ||::||:| ::||||:|| :::  |    :    |::|    :|: |::||:: :  :  |: |
Db          4 NKRVLILTANYGNGHVQVAKTLYEQCVRLGFQHVTV--SNLYQESNPIVSEVTQYLYLKS 61

Qy         65 FKYFRNMYKGFYYSRPDKLDKCFYK-----YYGL-NKLINLLIKE-KPDLILLTFPTPVM 117
              |  :  |: |||     :|| :  |       |: :  ||  :   |: | :||:|: |||  |:
Db         62 FSIGKQFYRLFYYG----VDKIYNKRKFNIYFKMGNKRLGELVDEHQPDIIINTFPMIVV 117

Qy        118 SVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGI 177
                   :   ||   ||||: ||| |:       :||||    |:  :::|  || ||:|||
Db        118 PEYRRRTGRVIPTFNVMTDFCLHKIWVHENVDKYYVATDYVKEKLLEIGTHPSNVKITGI 177
```

```
Qy     178 PIDNKFET-----PINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSANA 232
           ||  :||      || :|      || |:|: :|: ||| || |    :  : | |
Db     178 PIRPQFEESMPVGPIYKKY-----NLSPNKKVLLIMAGAHGVLKNVKELCEN-LVKDDQV 231

Qy     233 QVVMICGKSKELKRSLTA-------KFKLTRMYLILGYTKHMNEWMASSQLMITKPGGIT 285
           |||::|||:  || ||:|       | |      :||| : ::|    :  |||||||||
Db     232 QVVVVCGKNTALKESLSALEAENGDKLK------VLGYVERIDELFRITDCMITKPGGIT 285

Qy     286 ITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTNGNEQLTNM 345
           :||  |  :|:|   | ||||:|||  :||::|   : :  || :: | ||    : |  |
Db     286 LTEATAIGVPVILYKPVPGQEKENANFFEDRGAAIVVNRHEEILESVTSLLADEDTLHRM 345

Qy     346 ISTMEQDKIKYATQTICRDLL 366
           ::   :  ::: |  |:|
Db     346 KKNIKDLHLANSSEVILEDIL 366
```

```
RESULT   2
MURG_BORBU
ID   MURG_BORBU     STANDARD;      PRT;   363 AA.
AC   O51708;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG OR BB0767.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE001176; AAC67113.1; -.
DR   TIGR; BB0767; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   363 AA;  41118 MW;  1BAFA347384DB235 CRC64;
```

```
  Query Match             6.6%;  Score 136;  DB 1;  Length 363;
  Best Local Similarity   21.2%;  Pred. No. 0.0029;
  Matches   86;  Conservative   68;  Mismatches  155;  Indels   96;  Gaps   18;
```

```
Qy       5 NKKILIITGSFGNGHMQVTQSIVNQLNDMN-------LDHLSVIEHDLFMEAHPI--LTS 55
           ||||:  ||    ||:    ||: :| : :        :   : ||  |  |   | ::
Db       3 NKKIIFFTGGGTGGHVFPGISIIQKLKEFDNEIEFFWIGKKNSIEEKLIKEQDNIKFISI 62

Qy      56 ICKKW--YINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDLILLT-- 111
             | |   |  ||| | : :|                 |: |   :| | || ||  |
Db      63 PCGKLRRYF-SFKNFTDFFK---------------VILGIIKSFYVLKKYKPQLIYATGG 106

Qy     112 -FPTPV--------MSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDF 162
              ||          :  :| : :::  :|| :          : ::   ::: ||::: |
Db     107 FVSTPAIIASSLLKIKSITHEMDLDPGLATKIN---------SKFANNIHISFKESEKYF 157

Qy     163 IDVGIDPSTVKVTGIPIDNKFETPINQKQWLIDNNLDPDKQTILMSA--------GAFGV 214
             :       :  || ||  :|             |:|| : |              |
Db     158 KNY----KNIIYTGSPIRREF-------------LNPDPKIIKQLTQNTNKPIISILGG 199

Qy     215 SKGFDTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASS 274
           | | : :    |    :|:: |    | |  |: |  | | :    : | :  |
Db     200 SLGANALNNLALCIKKDAEIYFIHQSGKNL-NDLSEKNYLRRQFF---NAEEMASIVKFS 255

Qy     275 QLMITKPGGITITEGFAR---CIPMI-FLNPAPGQELENA---------FYFEEKGFGKI 321
           |:|:: |   | | ||    |  :| |   : | :::||          | :|    |
Db     256 NLIISRAGAGAIKE-FANAGACAILIPFKKGSRGDQIKNAKLLTNQNACIYIDEDEILNI 314

Qy     322 ADTPEEAIKIVASLTNGNEQLTNMISTMEQDKIKYATQTICRDLL 366
               :||:       |:: ::   :::   |::: | : |:
Db     315 -----NILKIIKKTLKDREKINSLKENIKKFNNKHSSTLIAKLLI 354
```

```
RESULT   3
MURG_ENTFA
ID   MURG_ENTFA     STANDARD;      PRT;   363 AA.
AC   O07109;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG.
OS   Enterococcus faecalis (Streptococcus faecalis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
OX   NCBI_TaxID=1351;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A24836;
RX   MEDLINE=97431524; PubMed=9287029;
RA   Pucci M.J., Thanassi J.A., Discotto L.F., Kessler R.E.,
RA   Dougherty T.J.;
RT   "Identification and characterization of cell wall-cell division gene
RT   clusters in pathogenic gram-positive cocci.";
RL   J. Bacteriol. 179:5632-5635(1997).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U94707; AAC45636.1; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
```

```
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   363 AA;  39916 MW;  5178BCC95264BE69 CRC64;


  Query Match             6.6%;  Score 135;  DB 1;  Length 363;
  Best Local Similarity   20.6%;  Pred. No. 0.0035;
  Matches   83;  Conservative   75;  Mismatches  161;  Indels   84;  Gaps   16;

Qy       9 LIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSFKY- 67
           :::||    ||:                 || :||          : |:     | |
Db       3 ILVTGGGTGGHIYPA-------------LSFVEH-----------VKKEAPATEFLYV 36

Qy      68 -----------------FRNM-YKGFYYSRPDKLDKCFYKYY-GLNKLINLLIKEKPDLI 108
                            |: :  :||  |   :  |  | :   :||   :: : :||::
Db      37 GTENGLESQIVPKAKIPFKTIKIQGFKRSLSPQNFKTIYLFLTSINKAKKIIREFQPDVV 96

Qy     109 LLTFPTPVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYY---------VATKETK 159
           : |      :|:     : ||  |:: :          : :||          ||:   |
Db      97 IGTGGYVSGAVVYAAHQLKIP--TIIHEQNSIPGMTNKFLSRYVDKIAICFPDVASFFPK 154

Qy     160 QDFIDVGIDPSTVKVTGIPIDNKFETPINQKQWLIDNNLDPDKQTILM---SAGAFGVSK 216
           :  |          :|| |   :  | : :   | :  ||| |:|:::   | ||  :::
Db     155 EKTI----------LTGNPRGQEVVT-VEKSAILSEFGLDPAKKTVVLFGGSRGALKINQ 203

Qy     217 GFDTMITDILAKSANAQVVMICGKS--KELKRSLTAKFKLTRMYLILGYTKHMNEWMASS 274
            |:      | :   ||:   |:   :||: ||    |      :  |   | | ||::
Db     204 AFEQAFP--LFEEREYQVLYASGERYYQELQESLKLSEKKLTNISVQPYIDKMVEVMANT 261

Qy     275 QLMITKPGGITITEGFARCIPMIFLNPAP----GQELENAFYFEEKG-FGKIADTPEEAI 329
           ||: : |  :| |  |  :| | : |+|       : :||    : |    | |
Db     262 DLMVGRAGATSIAEFTALGLPAILI-PSPYVTNDHQTKNAQSLVKVGAVEMIPDAELTGA 320

Qy     330 KIVASLTN---GNEQLTNMISTMEQDKIKYATQTICRDLLDLI 369
           ::||:: :    ||:   | :  : ::|  |:  : : :  |:
Db     321 RLVAAIDDILLNNEKRQQMATASKGERIPDASDRLYQVVKTLV 363


RESULT   4
N133_YEAST
ID   N133_YEAST     STANDARD;      PRT;  1157 AA.
AC   P36161;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   NUCLEOPORIN NUP133 (NUCLEAR PORE PROTEIN NUP133).
GN   NUP133 OR YKR082W OR YKR402.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JUXJR;
RX   MEDLINE=95112817; PubMed=7813444;
RA   Doye V., Wepf R., Hurt E.C.;
RT   "A novel nuclear pore protein Nup133p with distinct roles in poly(A)+
RT   RNA transport and nuclear pore distribution.";
RL   EMBO J. 13:6062-6075(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94262327; PubMed=8203164;
RA   Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,
RA   Remacha M.A., Revuelta J.L., Ballesta J.P.G., Jimenez A., del Rey F.;
RT   "The complete sequence of an 18,002 bp segment of Saccharomyces
RT   cerevisiae chromosome XI contains the HBS1, MRP-L20 and PRP16 genes,
RT   and six new open reading frames.";
RL   Yeast 10:231-245(1994).
CC   -!- FUNCTION: INVOLVED IN POLY(A)+ RNA TRANSPORT AND NUCLEAR PORE
CC       DISTRIBUTION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X80066; CAA56372.1; -.
DR   EMBL; Z27116; CAA81633.1; -.
DR   EMBL; Z28307; CAA82161.1; -.
DR   PIR; S38160; S38160.
DR   PIR; S39123; S39123.
DR   SGD; S0001790; NUP133.
KW   Nuclear protein; Transport; Transmembrane.
FT   TRANSMEM    217    233       POTENTIAL.
FT   SIMILAR      98    420       TO NUP120 (AA 434-763).
SQ   SEQUENCE   1157 AA;  133319 MW;  C8BDBB7D709C5C08 CRC64;


  Query Match             6.2%;  Score 127.5;  DB 1;  Length 1157;
  Best Local Similarity   20.9%;  Pred. No. 0.063;
  Matches  104;  Conservative   67;  Mismatches  147;  Indels  179;  Gaps   27;

Qy       2 VTQNKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIE--HDLF---------MEAH 50
             :| ||      | |    :  | |   :| |:|: : :::|   ||:          ::|
Db     285 ITTNK------GIFQTWQLSATNSHPTKLIDVNI-YEAILESLQDLYPFAHGTLKIWDSH 337

Qy      51 PI--------LTSI----CKK-WYINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLI 97
           |:        |:||    | : :|| |   | :   |              :  | ||  :
Db     338 PLQDESSQLFLSSIYDSSCNETYYILSTIIFDSSSNSF----------TIFSTYRLNTFM 387

Qy      98 NLL--IKEKPDL----------------ILLTFPTPVMSVLTEQFNINIPVATVMTDYRL 139
             :   | || :                ||: ||  |  |:|:   :|   : : : | :
Db     388 ESITDTKFKPKIFIPQMENANDTNEVTSILVMFPNAV--VITQ---VN---SKLDSSYSM 439

Qy     140 HKNWITPYSTR--------------YYVATKE--TKQDFI----------DVGIDPSTV- 172
           : |    | |                || ||:    | |:           :||   | |
Db     440 RRKWEDIVSLRNDIDIIGSGYDSKSLYVLTKQMGVLQFFVKENEETNSKPEVGFVKSHVD 499

Qy     173 ------KVTGIPIDNKFETPINQKQWLIDNNLDPDKQTILMSAGAF-----------GV 214
                 |:   |||       |:  |  |:::|    : |  | | :            |
Db     500 QAVYFSKINANPIDFNLPPEISLDQESIEHDLKLTSEEIFHSNGKYIPPMLNTLGQHLSV 559

Qy     215 SKGFDTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLTRMYL--------------- 259
             | |       :||: | ::      | |||  |  ||::    :
Db     560 RKEFFQNFLTFVAKNFNYKI------SPELKLDLIEKFEILNCCIKFNSIIRQSDVLNDI 613

Qy     260 ----ILGYTKHMNEWMASSQLMITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEE 315
               :  |    || :: ::|  |             |:||      ::  |   |
Db     614 WEKTLSNYNLTQNEHLTTKTVVINSPD----------VFPVIF------KQFLNHVVF-- 655

Qy     316 KGFGKIADTPEEAIKIVASLTNGNEQL--TNMISTMEQDKI-KYATQTICRDLLDLIGHS 372
                       ::    | | :| ·||:|:    | | :   :||  :||:|
Db     656 --------------VLFPSQNQNFKLNVTNLINLCFYDGILEEGEKTIRYELLEL---- 696

Qy     373 SQPQEI-YGKVPLYARF 388
             | |:   |:| :  |
Db     697 -DPMEVDTSKLPWFINF 712


RESULT   5
MURG_STRPN
ID   MURG_STRPN     STANDARD;      PRT;   352 AA.
AC   Q9ZHA9;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG.
OS   Streptococcus pneumoniae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
```

```
OC   Streptococcus.
OX   NCBI_TaxID=1313;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=G54;
RX   MEDLINE=99061199; PubMed=9846742;
RA   Massidda O., Anderluzzi D., Friedli L., Feger G.;
RT   "Unconventional organization of the division and cell wall gene
RT   cluster of Streptococcus pneumoniae.";
RL   Microbiology 144:3069-3078(1998).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF068902; AAC95450.1; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   352 AA;  39480 MW;  75AB4EC66B2DCFAD CRC64;

  Query Match             6.1%;  Score 124;  DB 1;  Length 352;
  Best Local Similarity   20.6%;  Pred. No. 0.023;
  Matches   80;  Conservative   80;  Mismatches  154;  Indels   74;  Gaps   19;

Qy       7 KILIITGSFGNGHMQVTQSI----------VNQLND-MNLDHLSVIEHDLFMEAHPILTS 55
Db       2 KKIVFTGGGTVGHVTLNLLLMPKFIEDGWEVHYIGDKRGIEHQEILKSGLDVTFHSIATG 61

Qy      56 ICKKWYINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDLI-----LL 110
Db      62 KLRR----------------YFSWQNMLD-VFKVCWGIVQSLFIMLRLRPQTLFSKGGFV 104

Qy     111 TFPTPVMSVLTEQFNININPVATVMTDYR--LHKNWITPYSTRYYVATKETKQDFIDVGID 168
Db     105 SVP-PVIAARVS----GVPVFIHESDLSMGLANKIAYKFATKMY-STFEQASSLSKVEHV 158

Qy     169 PSTVKVTGIPIDNKFETPINQKQWLID--NNLDPDKQTILMSAGAFGVSKGFDTMITDIL 226
Db     159 GAVTKVS----DQKNPEP----DELVDIQSHFNHKLPTVLFVGGSAG-ARVFNQLVTDHK 209

Qy     227 AK-SANAQVVMICGKS--KELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGG 283
Db     210 KELTERYNIINLTGDSSLNELSQNLFRVDYVTDLYQPL---------MELADIVVTRGGA 260

Qy     284 ITITEGFA----RCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTNGN 339
Db     261 NTIFELLAIAKLHVIVPLGREASRGDQLENAAYFVKKGYAE--DLQESDL----TLDSLE 314

Qy     340 EQLTNMISTMEQDKIKYATQTICRDLLD 367
Db     315 EKLSHLLSHKEDYQAKMKASKELKSLAD 342


RESULT   6
MURG_AQUAE
ID   MURG_AQUAE     STANDARD;      PRT;   344 AA.
AC   O67238;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG OR AQ_1177.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000727; AAC07193.1; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   344 AA;  39325 MW;  5A4B3322B4AC650C CRC64;

  Query Match             6.0%;  Score 123.5;  DB 1;  Length 344;
  Best Local Similarity   20.7%;  Pred. No. 0.024;
  Matches   80;  Conservative   72;  Mismatches  131;  Indels  103;  Gaps   19;

Qy       2 VTQNKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWY 61
Db      23 VLKKEKVKFIGSKRG---------IEYELKDL------IKTEKLFLDVEPLRE------- 60

Qy      62 INSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDLILL----TFPTPVM 117
Db      61 -------RNFYQ--------KL-KAIWKFLKAQEEINEFLKEDYRALIFGGYASLPLGIN 104

Qy     118 SVLTEQFNININPVATVMTDYRLHKNWITPYSTRYYVATKE---------TKQDFIDVGID 168
Db     105 TVLRRK------------ELFIHEQNSIPSKTNKILSKKAKKVLITFNYTKRFF------ 146

Qy     169 PSTVKVTGIPIDNKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAK 228
Db     147 PEGVRV-GLPIRKELKKKLPKKEVKKRFGLEPDKITVLI----FGGSQG--ALFLNELAR 199

Qy     229 SANA------QVVMICGK--SKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITK 280
Db     200 DLKSVLPKEFQVILLTGKIHYEKFKNLEGEKFR------VMPFSLDMGLIYSASDVAISR 253

Qy     281 PGGITITEGFARCIPMIFLNPAPGQELENAFY----FEEKGFGKIADTPE-------EAI 329
Db     254 AGAGTINELSHFGVPSVFV-PYPYAVDDHQFYNAKEIEKLGGGLVLRQEEAKPDKVLSAL 312

Qy     330 K-IVASLTNGNEQLTNMISTMEQDKI 354
Db     313 KEIVKNLERYSENIKKFFAEGAEERM 338


RESULT   7
```

```
SR54_MYCMY
ID   SR54_MYCMY      STANDARD;      PRT;   447 AA.
AC   Q01442;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   SIGNAL RECOGNITION PARTICLE PROTEIN (FIFTY-FOUR HOMOLOG).
GN   FFH OR SRPM54.
OS   Mycoplasma mycoides.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Entomoplasmataceae.
OX   NCBI_TaxID=2102;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93087189; PubMed=1280809;
RA   Samuelsson T.B.;
RT   "A Mycoplasma protein homologous to mammalian SRP54 recognizes a
RT   highly conserved domain of SRP RNA.";
RL   Nucleic Acids Res. 20:5763-5770(1992).
CC   -!- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC       PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC       RIBOSOMES (BY SIMILARITY).
CC   -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 4.5 RNA
CC       MOLECULE AND PROTEIN FFH (BY SIMILARITY).
CC   -!- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE: THE G-DOMAIN
CC       BINDS GTP; THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC       SEQUENCE.
CC   -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M91593; AAA25441.1; -.
DR   PIR; S27591; S27591.
DR   PIR; S35481; S35481.
DR   HSSP; O07347; 1FFH.
DR   InterPro; IPR000897; -.
DR   Pfam; PF00448; SRP54; 1.
DR   PROSITE; PS00300; SRP54; 1.
KW   Signal recognition particle; GTP-binding; RNA-binding.
FT   DOMAIN        1    295       G-DOMAIN.
FT   DOMAIN      296    447       M-DOMAIN.
FT   NP_BIND     108    115       GTP (BY SIMILARITY).
FT   NP_BIND     190    194       GTP (BY SIMILARITY).
FT   NP_BIND     248    251       GTP (BY SIMILARITY).
SQ   SEQUENCE   447 AA;  50100 MW;  24A71128D4041D43 CRC64;


  Query Match             6.0%;  Score 122;  DB 1;  Length 447;
  Best Local Similarity   22.9%;  Pred. No. 0.045;
  Matches   80;  Conservative   52;  Mismatches  143;  Indels   74;  Gaps   16;

Qy      13 GSFGNGHMQ--VTQSIVNQ-LNDMNLD------HLSVIEHDLFMEAHPILTSICKKWYIN 63
           | | :  ||  : ::: |  ||: |:        ||::| |: :||         |  ||
Db       4 GDFLSKRMQKSIEKNMKNSTLNEENIKETLKEIRLSLLEADVNIEA--------AKEIIN 55

Qy      64 SFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDLILLTFPTPVMSVLTEQ 123
           : |      | | |     :   |     :|:|:| ||   | :   |: || |  :
Db      56 NVK---QKALGGYISEGASAHQQMIKIVH-EELVNILGKENAPLDINKKPSVVMMVGLQ- 110

Qy     124 FNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGID---PSTVKVTGIPID 180
                               |  :   |:  |: |:  : ||:|   |  ::   : :
Db     111 ---------------GSGKTTTANKLAYLLNKKNKKKVLLVGLDIYRPGAIEQL-VQLG 153

Qy     181 NKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSANAQVVMICGK 240
           |  | : :|       || |      |  :     || :| |   :    ||:|
Db     154 QKTNTQVFEK-----GKQDPVKTA--EQALEYAKENNFDVVILDTAGRLQVDQVLM---- 202

Qy     241 SKELKRSLTAKFKLTRMYL---------ILGYTKHMNEWMASSQLMITK-----PGGITI 286
           |||  :|  |      : |          |:  |   |: :  | :::||      || |:
Db     203 -KELD-NLKKKTSPNEILLVVDGMSGQEIINVTNEFNDKLKLSGVVVTKLDGDARGGATL 260

Qy     287 TEGFARCIPMIFLNPAPGQELENAFYFEE-----KGFGKIADTPEEAIK 330
           :  :   :|: |:     |      ||| :       | | |    | |::
Db     261 SISYLTKLPIKFIGEGEGYNALAAFYPKRMADRLMGMGDIETLFERAVE 309


RESULT   8
UD17_RAT
ID   UD17_RAT       STANDARD;      PRT;   531 AA.
AC   Q64633;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   UDP-GLUCURONOSYLTRANSFERASE 1-7 PRECURSOR, MICROSOMAL (EC 2.4.1.17)
DE   (UDPGT) (UGT1*7) (UGT1-07) (UGT1.7) (UGT1A7) (A2).
GN   UGT1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-286 FROM N.A.
RC   STRAIN=WISTAR;
RX   MEDLINE=95332265; PubMed=7608130;
RA   Emi Y., Ikushiro S.I., Iyanagi T.;
RT   "Drug-responsive and tissue-specific alternative expression of
RT   multiple first exons in rat UDP-glucuronosyltransferase family 1
RT   (UGT1) gene complex.";
RL   J. Biochem. 117:392-399(1995).
RN   [2]
RP   SEQUENCE OF 287-531 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=90274676; PubMed=2112380;
RA   Sato H., Koiwai O., Tanabe K., Kashiwamata S.;
RT   "Isolation and sequencing of rat liver bilirubin UDP-
RT   glucuronosyltransferase cDNA: possible alternate splicing of a common
RT   primary transcript.";
RL   Biochem. Biophys. Res. Commun. 169:260-264(1990).
CC   -!- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
CC       SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
CC       ENDOGENOUS COMPOUNDS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCURONATE + ACCEPTOR = UDP + ACCEPTOR
CC       BETA-D-GLUCORONOSIDE.
CC   -!- SUBCELLULAR LOCATION: MICROSOMAL.
CC   -!- ALTERNATIVE PRODUCTS: THE UGT1 GENE ENCODES FOR MANY DIFFERENT
CC       UDPGT ISOZYMES WHICH HAVE A DIFFERENT N-TERMINAL DOMAIN AND A
CC       COMMON C-TERMINAL DOMAIN OF 245 RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D38062; BAA07258.1; -.
DR   EMBL; M34007; AAA42312.1; ALT_TERM.
DR   InterPro; IPR002213; -.
DR   Pfam; PF00201; UDPGT; 1.
DR   PROSITE; PS00375; UDPGT; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
KW   Multigene family; Microsome; Alternative splicing.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    531       UDP-GLUCURONOSYLTRANSFERASE 1-7.
FT   TRANSMEM    487    503       POTENTIAL.
FT   CARBOHYD     71     71       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    431    431       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
SQ   SEQUENCE   531 AA;  59627 MW;  BC791DCE724CA621 CRC64;

  Query Match             5.8%;  Score 119;  DB 1;  Length 531;
  Best Local Similarity   18.3%;  Pred. No. 0.096;
  Matches   79;  Conservative   70;  Mismatches  123;  Indels  160;  Gaps   20;

Qy      23 TQSIVNQLNDMNLDHLSVIEHDLFMEAH----------PILTSICKKWYINSFKYFRNMY 72
           | |: :   |:|       |   |:::           |:|||  : ::   | : |:::
Db      76 TYSVSHTQEDLNR------EFKFFIDSQWKTQQESGVLPLLTSPAQGFFELLFSHCRSLF 129

Qy      73 KG---FYYSRPDKLDKCFY----------KYYGLNKLI---------------------- 97
           |     | :    |  |             ||: |  ::
Db     130 KDKKLVEYLKQSSFDAVFLDPFDVCGLTVAKYFSLPSVVFSRGIFCHYLEEGSQCPSPPS 189

Qy      98 -------------------NLL-----------IKEKPDLILLTFPTPVMSVLTEQFN 125
                              |||           |   |:      |||   :|: |:
Db     190 YVPRPILKLTDTMTFKERVWNLLSYMGEHAFCPSFFKTATDIASEVLQTPV--TMTDLFS 247

Qy     126 INIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGI------PI 179
             : |  : ||: |                        :   : |: : : ||      |:
Db     248 -PVSVWLLRTDFTLE---------------------LPRPVMPNVIHIGGINCHQRKPV 284

Qy     180 DNKFETPINQKQWLIDNNLDPDKQTILMSAGAFGV---SKGFDTMITDILAKSANAQVVM 236
           :||  :|              ::|  |:   | |  :|:::|  | |  ::
Db     285 SKEFEAYVN-----------------ASGEHGIVVFSLG--SMVSEIPEKKA-MEIAE 322

Qy     237 ICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLM--------ITKPGGITITE 288
           |:   : ::|  ::  ||   :   |  : :|:  : |:          ||  |   | |
Db     323 ALGR---IPQTLLWRYTGTRPSNLAKNT-ILVKWLPQNDLLGHPKARAFITHSGSHGIYE 378

Qy     289 GFARCIPMIFLNPAPGQELENAFYFEEKGFG-------KIADTPEEAIKIVASLTNGNEQ 341
           |    :||: : |  | :::||   | :| |          ||  | |:| | :  :  |
Db     379 GICNGVPMVMM-PLFGDQMDNAKRMETRGAGVTLNVLEMTADDLENALKTVINNKSYKEN 437

Qy     342 LTNMISTMEQDK 353
           :   :|:: :|:
Db     438 IMR-LSSLHKDR 448


RESULT   9
MURG_HELPY
ID   MURG_HELPY     STANDARD;      PRT;   353 AA.
AC   O25770;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG OR HP1155.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000621; AAD08196.1; -.
DR   TIGR; HP1155; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   353 AA;  39474 MW;  6D136CD92A180513 CRC64;

  Query Match             5.7%;  Score 117.5;  DB 1;  Length 353;
  Best Local Similarity   20.6%;  Pred. No. 0.072;
  Matches   81;  Conservative   59;  Mismatches  169;  Indels   85;  Gaps   20;

Qy      11 ITGSFGNGHMQVTQSIVNQLNDMNLDHL---SVIEHDLFMEAHPILTSICKKWYINSFKY 67
           :||    ||: : :::  :|    :: :   |    |              |:|: ||
Db       5 LTGGGTGGHLSIAKALAIELEKQGIEAIYLGSTYGQD------------KEWFENS-PL 50

Qy      68 FRNMYKGFYYSRPDKLDKCFYKYYG---LNKLINLLIKE--KPDLILLTF--------PT 114
           |   |   :::    ::| |:|  |   |         ||  |   | |          |
Db      51 FSERY---FFNTQGVVNKSFFKKIGSLFLQAKAAFKAKEILKKHQITHTISVGGFSAGPA 107

Qy     115 PVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGI-DPSTVK 173
             |:| :     ||:      |   :| |     ||   : : |  |       |
Db     108 SFASLLNK-----IPL------YIHEQNAIKGSLNRY--LSPKAKAVFSSYAFKDKGNHV 154

Qy     174 VTGIPIDNKF------ETPINQKQWLIDNNLDPDKQTILMSAGAFGVS--KGFDTMITDI 225
           :|  |: | |       | |                  ||   |: |      |  :
Db     155 LTSYPVQNAFFDFARTRTEIKH--------------ILFLGGSQGAKAINEFALLNAPK 199

Qy     226 LAKSANAQVVMICG-KSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGI 284
           | |    ::  |||  | |  |    :  |     :  :  :: | |  : | ::: |
Db     200 LTKQ-GIKITHICGPNSYEQVRFFYQELGLLDKIELFAFHNNITEIMHRADLCVSRAGAS 258

Qy     285 TITEGFARCIPMIFLNPAP----GQELENAFYFEEKGFGKIAD----TPEEAIKIVASLT 336
           :: |  |  :| ||: | |      :  |   ||::    :        |::  :::  |
Db     259 SVWELCANGLPTIFI-PYPFASNNHQYYNVLEFEKENLCYVVPQNELLPKKLFEVIRKLN 317

Qy     337 ----NGNEQLTNMISTMEQDKIKYATQTICRDLL 366
                ||: || : : ::|   |   :||   :|
Db     318 QKDDQGNKNLTTISNQLQQKIAKDGAKTIIETIL 351


RESULT  10
MURG_HELPJ
ID   MURG_HELPJ     STANDARD;      PRT;   353 AA.
AC   Q9ZK59;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG OR JHP1082.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic-sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001535; AAD06652.1; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   353 AA;  39506 MW;  1773AD7E5C214EF9 CRC64;

  Query Match             5.7%;  Score 116.5;  DB 1;  Length 353;
  Best Local Similarity   19.8%;  Pred. No. 0.086;
  Matches   79;  Conservative   63;  Mismatches  162;  Indels   95;  Gaps   20;

Qy      11 ITGSFGNGHMQVTQSIVNQLNDMNLDHL---SVIEHDLFMEAHPILTSICKKWYINSFKY 67
           :||    ||: : :::  :|    :: :   |    |              |:|: ||
Db       5 LTGGGTGGHLSIAKALAIELEKQGIEAIYLGSTYGQD------------KEWFENS-PL 50

Qy      68 FRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDL----------ILLTF----- 112
           |   |   :::    ::| |:|     |: :| :: |             |  |
Db      51 FSERY---FFNTQGVVNKSFFK-----KIRSLFLQAKAAFKAKEILKKHQITHTISVGGF 102

Qy     113 ---PTPVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGI-D 168
              |    |:| :     ||:      |   :| |     ||   : : |  |      |
Db     103 SAGPASFASLLNK-----IPL------YIHEQNAIKGSLNRY--LSPKAKAVFSSYAFKD 149

Qy     169 PSTVKVTGIPIDNKF------ETPINQKQWLIDNNLDPDKQTILMSAGAFGVS--KGFDT 220
               :|  |: | |       | |                 ||   |: |      |
Db     150 KGNHVLTSYPVQNAFFDHARTRTEIKH--------------ILFLGGSQGAKAINEFAL 194

Qy     221 MITDILAKSANAQVVMICG-KSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMIT 279
           :    | |    ::  |||  | |  |    :  |     :  :  :: | |  : | ::
Db     195 LNAPKLTKQ-GIKITHICGPNSYEQVRFFYQELGLLDKIELFAFHNNITEVMHRADLCVS 253

Qy     280 KPGGITITEGFARCIPMIFLNPAP----GQELENAFYFEEKGFGKIAD----TPEEAIKI 331
           : |  :: |  |  :| ||: | |      :  |   ||::    :         |::  ::
Db     254 RAGASSVWELCANGLPTIFI-PYPFASNNHQYYNVLEFEKENLCYVVPQNELLPKKLFEV 312

Qy     332 VASLT----NGNEQLTNMISTMEQDKIKYATQTICRDLL 366
           :  |      ||: || : : ::|   |   :||   :|
Db     313 IRKLNQKDDQGNKNLTTISAKLQQKIAKDGAKTIIETIL 351


RESULT  11
Y511_RICPR
ID   Y511_RICPR     STANDARD;      PRT;   950 AA.
AC   Q9ZD36;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PROTEIN RP511.
GN   RP511.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ235272; CAA14963.1; -.
DR   InterPro; IPR001646; -.
DR   Pfam; PF00805; Pentapeptide; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   950 AA;  108612 MW;  C047F8BCF063F715 CRC64;

  Query Match             5.5%;  Score 113;  DB 1;  Length 950;
  Best Local Similarity   20.4%;  Pred. No. 0.61;
  Matches   92;  Conservative   69;  Mismatches  156;  Indels  134;  Gaps   21;

Qy       4 QNKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIE--HDLFMEAHPILTSICKKWY 61
             : ||:| |     | |  | | :|| :|: |     :::  : || :   :|||  |:
Db     209 EKKKLLKIF----NTHPDVKQELVNNINNPN-----ILKKFNKLFYKQEIMLTSFLKEVK 259

Qy      62 INSFKYFRNMYKGF-------------------------YYSRPDK-----LDKCFYKYY 91
             |  : :  :: :                          : |:|     |:|
Db     260 AQSKPFLQEHFESYKIDLKILDIIPTLLNKIPDIKEIFDTLNAPNKGVMISLEKALEMVA 319

Qy      92 GLNKLINLLIKEK---PDLIL-LTFPTPVMSVLTEQFNIN----IPVATVMTDYRLHKNW 143
           | ::| :     |   |:: | :   || :  :| ::| :    : |  ||:   :
Db     320 GDDQLKSFFANNKTILPNMALGIIENTPSVQSITNEYNFDQQMLVIVGEVMSKPEIAHEI 379

Qy     144 ITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGIPIDNKFETPINQKQWLIDNNLD---- 199
           |   :   |::      : | : |||  |:  | ::         |: | || ::
Db     380 IADLNKGDYMS---LTGNIISIINDPS-FKLKDILVE-------QSKKGLFDNLINGVLE 428

Qy     200 ---PDKQTILMSAGAFGVSKGFDTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLTR 256
             : | |      :|:  |  | :| |:        ::  | : ||       |: |
Db     429 QDIKNSQIIKQQLINYGMEAGDVTKLTKIMP--------ILLDKPESLK-------KVFR 473

Qy     257 MYLILGYTKHMNEWMASSQLMITKPGG------------ITITEGFARCIPMIFLNPAP 303
             ::   |||      |    : :||              :| :      || |  |
Db     474 DFIKGNYTK-----MTKELISLTKDNPKIKEYLNNNRAIFASILDKTLMDIPGI--NNLD 526

Qy     304 GQELENAFYFEEKGFGKIADTPEEAIKIV-----------------ASLTNGNE-QLTN 344
           |||  |           : : |:| ||::                  |  ||  | || |
Db     527 KQELYNI-------LPSMLNHPDELIKVIEEVEKSHYHGAVSAIYNLAQKTNYFEGQLPN 579

Qy     345 MISTMEQDKIKYATQTI------CRDLLDLI 369
           :|         |||: :       ||  | :
Db     580 IIKAGFNSGFNYATEKVKDVFSSSRDFKDKV 610


RESULT  12
MURG_ENTHR
ID   MURG_ENTHR     STANDARD;      PRT;   360 AA.
AC   O07670;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
```

```
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG.
OS   Enterococcus hirae.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC   Enterococcus.
OX   NCBI_TaxID=1354;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S185;
RX   MEDLINE=99449055; PubMed=10520745;
RA   Duez C., Thamm I., Sapunaric F., Coyette J., Ghuysen J.-M.;
RT   "The division and cell wall gene cluster of Enterococcus hirae S185.";
RL   DNA Seq. 9:149-161(1998).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y13922; CAA74235.1; -.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
SQ   SEQUENCE   360 AA;  39591 MW;  444D6F7CFEC1FDA1 CRC64;
```

```
  Query Match             5.4%;  Score 111.5;  DB 1;  Length 360;
  Best Local Similarity   19.5%;  Pred. No. 0.21;
  Matches   76;  Conservative   81;  Mismatches  170;  Indels   63;  Gaps   17;

Qy       9 LIITGSFGNGHMQVTQSIVN--QLNDMNLDHLSV-----IEHDLFME-AHPILT----SI 56
           :::||    ||:    : ||  :  : | : : |     :|: :  |   |  |
Db       3 ILVTGGGTGGHIYPALAFVNYVKTKEPNTEFMYVGAQRGLENKIVPETGMPFRTLEIQGF 62

Qy      57 CKKWYINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKE-KPDLILLTFPTP 115
             :|  ::: |  :   |    ::                    ::|| |||::: |
Db      63 QRKLSLHNLKTIQLFLKSIREAKK------------------ILKEFKPDVVIGTGGYV 103

Qy     116 VMSVLTEQFNINIPVATVMTDYR----LHKNWITPYSTRYYVATKETKQDFIDVGIDPST 171
             :|:     : ||  |:: :      :   ::: |  |  :: ::    |
Db     104 SGAVVYAASKLAIP--TIIHEQNSVPGITNKFLSRYVDRIALSFEDAAPFF-----PAEK 156

Qy     172 VKVTGIPIDNKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSAN 231
             : | |   :    ::: : |    |||:|:|:|:  |: |  |  :  :|: |  | :
Db     157 SSLIGNPRAQEV-ADMDKSKILATYGLDPEKKTVLIFGGSQGALK-INQAVTEFL-MSFD 213

Qy     232 AQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITITEGFA 291
             :  ::    :   : :  |         |  |    | | |||| |:: : |  :| |  |
Db     214 QEYQVLYASGERYYKDIQTKVPACANVSIQPYINKMAEVMASSDLLVGRAGATSIAELTA 273

Qy     292 RCIPMIFLNPAP----GQELENAFYFEEKGFGKIADTPE-------EAIKIVASLTNGNE 340
             :| | : |:|      : :||    :    |:    |       :||: : :    |:
Db     274 LGLPAILI-PSPYVTNDHQTKNAMSLVKNNAAKMIKDDELDGRSLKQAIEEIMT----ND 328

Qy     341 QLTNMIS-TMEQDKIKYATQTICRDLLDLI 369
           ||   :|   :|  |  |:: :   :  ||
Db     329 QLQKQMSLASKQQGIPDASERMYELVKSLI 358

RESULT  13
```

```
MURG_BACSU
ID   MURG_BACSU     STANDARD;      PRT;   363 AA.
AC   P37585; P18578; Q59247;
DT   01-NOV-1990 (Rel. 16, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--N-ACETYLMURAMYL-(PENTAPEPTIDE)
DE   PYROPHOSPHORYL-UNDECAPRENOL N-ACETYLGLUCOSAMINE TRANSFERASE
DE   (EC 2.4.1.-) (UNDECAPRENYL-PP-MURNAC-PENTAPEPTIDE-UDPGLCNAC GLCNAC
DE   TRANSFERASE).
GN   MURG.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92380484; PubMed=1387377;
RA   Miyao A., Yoshimura A., Sato T., Yamamoto T., Theeragool G.,
RA   Kobayashi Y.;
RT   "Sequence of the Bacillus subtilis homolog of the Escherichia coli
RT   cell-division gene murG.";
RL   Gene 118:147-148(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=93003529; PubMed=1391053;
RA   Henriques A.O., de Lencastre H., Piggot P.J.;
RT   "A Bacillus subtilis morphogene cluster that includes spoVE is
RT   homologous to the mra region of Escherichia coli.";
RL   Biochimie 74:735-748(1992).
RN   [3]
RP   SEQUENCE OF 216-363 FROM N.A.
RX   MEDLINE=90078133; PubMed=2556375;
RA   Beall B., Lutkenhaus J.;
RT   "Nucleotide sequence and insertional inactivation of a Bacillus
RT   subtilis gene that affects cell division, sporulation, and
RT   temperature sensitivity.";
RL   J. Bacteriol. 171:6821-6834(1989).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II).
CC   -!- PATHWAY: LAST STEP OF PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10602; BAA01454.1; -.
DR   EMBL; X64259; CAA45558.1; -.
DR   EMBL; M31827; AAA83968.1; -.
DR   EMBL; Z99111; CAB13395.1; -.
DR   PIR; JC1275; JC1275.
DR   PIR; S26499; S26499.
DR   SubtiList; BG10227; murG.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall;
KW   Membrane; Peptidoglycan synthesis.
FT   CONFLICT    181    181       K -> G (IN REF. 1).
FT   CONFLICT    270    270       A -> T (IN REF. 1).
SQ   SEQUENCE   363 AA;  39977 MW;  185B20688C732489 CRC64;
```

```
  Query Match             5.4%;  Score 111.5;  DB 1;  Length 363;
  Best Local Similarity   24.1%;  Pred. No. 0.21;
  Matches   55;  Conservative   46;  Mismatches   92;  Indels   35;  Gaps   11;
```

```
Qy      92 GLNKLINLLIKEKPDLILLTFPTPVMSVLTEQFNINIPVATVMTDYR----LHKNWITPY 147
           |: |  : | : ||| :: |       |:     : ||  |:: :      :   ::: |
Db      80 GVKKSKSYLAEFKPDAVIGTGGYVCGPVVYAAAKMGIP--TIVHEQNSLPGITNKFLSKY 137

Qy     148 STRYYVATKETKQDFIDVGIDPS-TVKVTGIPIDNKFETPINQKQWLIDNNLDPDKQTIL 206
           :  :  :| | |      || |  || |  ::    |  : | :  |  ||:|:|
Db     138 VNKVAICFEEAKSHF------PSEKVVFTGNPRASEV-VSIKTGRSLAEFKLSEDKKTVL 190

Qy     207 MSAGAFGVSKG---FDTMITDI--LAKSANAQVVMICGK------SKELKRSLTAKFKLT 255
           :    || |:|    :  : |:  : |: : ||: | |:        |||    |   :|
Db     191 I----FGGSRGAAPINRAVIDMQDVLKTRDYQVLYITGEVHYEKVMNELKSKGAADNMVT 246

Qy     256 RMYLILGYTKHMNEWMASSQLMITKPGGITITEGFARCIPMIFLNPAP 303
           : :|       | |:: :  ::: : |  || |  |  || : : |:|
Db     247 KPFL-----HQMPEYLKAIDVIVARAGAATIAEITALGIPSVLI-PSP 288


RESULT 14
UD14_HUMAN
ID   UD14_HUMAN     STANDARD;      PRT;   534 AA.
AC   P22310;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   UDP-GLUCURONOSYLTRANSFERASE 1-4 PRECURSOR, MICROSOMAL (EC 2.4.1.17)
DE   (UDPGT) (UGT-1D) (UGT1*4) (UGT1-04) (UGT1.4) (UGT1A4) (UGT1D)
DE   (BILIRUBIN SPECIFIC ISOZYME 2) (HUG-BR2).
GN   UGT1 OR GNT1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=91093210; PubMed=1898728;
RA   Ritter J.K., Crawford J.M., Owens I.S.;
RT   "Cloning of two human liver bilirubin UDP-glucuronosyltransferase
RT   cDNAs with expression in COS-1 cells.";
RL   J. Biol. Chem. 266:1043-1047(1991).
RN   [2]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX   MEDLINE=92147680; PubMed=1339448;
RA   Ritter J.K., Chen F., Sheen Y.Y., Tran H.M., Kimura S., Yeatman M.T.,
RA   Owens I.S.;
RT   "A novel complex locus UGT1 encodes human bilirubin, phenol, and
RT   other UDP-glucuronosyltransferase isozymes with identical carboxyl
RT   termini.";
RL   J. Biol. Chem. 267:3257-3261(1992).
RN   [3]
RP   VARIANT CN-I PHE-376.
RX   MEDLINE=92339803; PubMed=1634050;
RA   Bosma P.J., Chowdhury J.R., Huang T.-J., Lahiri P., Elferink R.P.J.O.,
RA   van Es H.H.G., Lederstein M., Whitington P.F., Jansen P.L.,
RA   Chowdhury N.R.;
RT   "Mechanisms of inherited deficiencies of multiple UDP-
RT   glucuronosyltransferase isoforms in two patients with Crigler-Najjar
RT   syndrome, type I.";
RL   FASEB J. 6:2859-2863(1992).
RN   [4]
RP   VARIANTS CN-II PRO-132 AND ASP-487.
RX   MEDLINE=94107323; PubMed=8280139;
RA   Aono S., Yamada Y., Keino H., Hanada N., Nakagawa T., Sasaoka Y.,
RA   Yazawa T., Sato H., Koiwai O.;
RT   "Identification of defect in the genes for bilirubin UDP-glucuronosyl-
RT   transferase in a patient with Crigler-Najjar syndrome type II.";
RL   Biochem. Biophys. Res. Commun. 197:1239-1244(1993).
RN   [5]
RP   VARIANT CN-II ARG-332.
RX   MEDLINE=94102756; PubMed=8276413;
RA   Moghrabi N., Clarke D.J., Boxer M., Burchell B.;
RT   "Identification of an A-to-G missense mutation in exon 2 of the UGT1
RT   gene complex that causes Crigler-Najjar syndrome type 2.";
RL   Genomics 18:171-173(1993).
CC   -!- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
CC       SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
CC       ENDOGENOUS COMPOUNDS. GLUCURONIDATES BILIRUBIN IX-ALPHA TO FORM
CC       BOTH THE IX-ALPHA-C8 AND IX-ALPHA-C12 MONOCONJUGATES AND
CC       DICONJUGATE.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCURONATE + ACCEPTOR = UDP + ACCEPTOR
CC       BETA-D-GLUCORONOSIDE.
CC   -!- SUBCELLULAR LOCATION: MICROSOMAL.
CC   -!- ALTERNATIVE PRODUCTS: THE UGT1 GENE ENCODES FOR MANY DIFFERENT
CC       UDPGT ISOZYMES WHICH HAVE A DIFFERENT N-TERMINAL DOMAIN AND A
CC       COMMON C-TERMINAL DOMAIN OF 245 RESIDUES.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER. NOT EXPRESSED IN SKIN OR
CC       KIDNEY.
CC   -!- INDUCTION: BY PHENOBARBITAL.
CC   -!- DISEASE: THE GILBERT'S SYNDROME IS SHOWN TO OCCUR AS A CONSEQUENCE
CC       OF REDUCED BILIRUBIN TRANSFERASE ACTIVITY. THE DISORDER, IS MOST
CC       OFTEN DETECTED IN YOUNG ADULTS WITH VAGUE NONSPECIFIC COMPLAINTS.
CC       A MORE SEVERE INHERITABLE DEFICIENCY IN BILIRUBIN ACTIVITY EXIST
CC       IN CRIGLER-NAJJAR (CN): PATIENTS WITH TYPE I (RECESSIVE TRAIT)
CC       HAVE SEVERE HYPERBILIRUBINEMIA AND USUALLY DIE OF KERNICTERUS
CC       (BILIRUBIN ACCUMULATION IN THE BASAL GANGLIA AND BRAINSTEM NUCLEI)
CC       WITHIN THE FIRST YEAR OF LIFE. PATIENTS WITH TYPE II (DOMINANT
CC       TRAIT) HAVE LESS SEVERE HYPERBILIRUBINEMIA AND USUALLY SURVIVE
CC       INTO ADULTHOOD WITHOUT NEUROLOGIC DAMAGE. PHENOBARBITAL, WHICH
CC       INDUCES THE PARTIALLY DEFICIENT GLUCURONYL TRANSFERASE, CAN
CC       DIMINISH THE JAUNDICE.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M84128; AAA61249.1; -.
DR   EMBL; M84124; AAA61247.1; ALT_SEQ.
DR   EMBL; M84122; AAA61247.1; JOINED.
DR   EMBL; M84123; AAA61247.1; JOINED.
DR   EMBL; M57951; AAA63196.1; -.
DR   PIR; B39092; B39092.
DR   MIM; 191740; -.
DR   MIM; 143500; -.
DR   MIM; 218800; -.
DR   InterPro; IPR002213; -.
DR   Pfam; PF00201; UDPGT; 1.
DR   PROSITE; PS00375; UDPGT; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
KW   Multigene family; Microsome; Alternative splicing; Disease mutation.
FT   SIGNAL       1     28       POTENTIAL.
FT   CHAIN       29    534       UDP-GLUCURONOSYLTRANSFERASE 1-4.
FT   TRANSMEM    492   508       POTENTIAL.
FT   CARBOHYD    119   119       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    142   142       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    296   296       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    348   348       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     132   132       L -> P (IN CRIGLER-NAJJAR TYPE II).
FT                               /FTId=VAR_009506.
FT   VARIANT     332   332       Q -> R (IN CRIGLER-NAJJAR TYPE II).
FT                               /FTId=VAR_007710.
FT   VARIANT     376   376       S -> F (IN CRIGLER-NAJJAR TYPE I).
FT                               /FTId=VAR_007711.
FT   VARIANT     487   487       Y -> D (IN CRIGLER-NAJJAR TYPE II).
FT                               /FTId=VAR_009507.
SQ   SEQUENCE   534 AA;  60025 MW;  AB745D46F57538BE CRC64;


 Query Match             5.4%;  Score 111.5;  DB 1;  Length 534;
 Best Local Similarity   20.8%;  Pred. No. 0.36;
 Matches   60;  Conservative   54;  Mismatches  103;  Indels   71;  Gaps   12;
```

```
Qy      93 LNKLINLLIKEKPDLILLTFPTPVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYY 152
           | :: |:|       |  ||  |  |: :| |   : |  :::                |
Db     208 LQRVKNMLYPLALSYICHTFSAPYASLASELFQREVSVVDLVS----------------Y 251

Qy     153 VATKETKQDFI---DVGIDPSTVKVTGI------PIDNKFETPINQKQWLIDNNLDPDKQ 203
           :    : ||:      | |: | : ||      |:  :||  ||
Db     252 ASVWLFRGDFVMDYPRPIMPNMVFIGGINCANGKPLSQEFEAYIN--------------- 296

Qy     204 TILMSAGAFGV---SKGFDTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLTRMYLI 260
             ::|  |:   | |  :|:::|  | | |    |     :: :::  ::  ||   :
Db     297 ----ASGEHGIVVFSLG--SMVSEIPEKKAMA----IADALGKIPQTVLWRYTGTRPSNL 346

Qy     261 LGYTKHMNEWMASSQLM--------ITKPGGITITEGFARCIPMIFLNPAPGQELENAFY 312
             |  : :|:  : |:         ||  |   : |     :||: : |  | :::||
Db     347 ANNT-ILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMM-PLFGDQMDNAKR 404

Qy     313 FEEKGFG-------KIADTPEEAIKIVASLTNGNEQLTNMISTMEQDK 353
           | || |         ::  | |:| | :  :  | :   :|:: :|:
Db     405 METKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMR-LSSLHKDR 451


RESULT 15
UD15_RAT
ID   UD15_RAT       STANDARD;      PRT;   531 AA.
AC   Q64638;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   UDP-GLUCURONOSYLTRANSFERASE 1-5 PRECURSOR, MICROSOMAL (EC 2.4.1.17)
DE   (UDPGT) (UGT1*5) (UGT1-05) (UGT1.5) (UGT1A5) (B5).
GN   UGT1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-286 FROM N.A.
RC   STRAIN=WISTAR;
RX   MEDLINE=95332265; PubMed=7608130;
RA   Emi Y., Ikushiro S.I., Iyanagi T.;
RT   "Drug-responsive and tissue-specific alternative expression of
RT   multiple first exons in rat UDP-glucuronosyltransferase family 1
RT   (UGT1) gene complex.";
RL   J. Biochem. 117:392-399(1995).
RN   [2]
RP   SEQUENCE OF 287-531 FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=90274676; PubMed=2112380;
RA   Sato H., Koiwai O., Tanabe K., Kashiwamata S.;
RT   "Isolation and sequencing of rat liver bilirubin UDP-
RT   glucuronosyltransferase cDNA: possible alternate splicing of a common
RT   primary transcript.";
RL   Biochem. Biophys. Res. Commun. 169:260-264(1990).
CC   -!- FUNCTION: UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
CC       SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
CC       ENDOGENOUS COMPOUNDS.
CC   -!- CATALYTIC ACTIVITY: UDP-GLUCURONATE + ACCEPTOR = UDP + ACCEPTOR
CC       BETA-D-GLUCORONOSIDE.
CC   -!- SUBCELLULAR LOCATION: MICROSOMAL.
CC   -!- ALTERNATIVE PRODUCTS: THE UGT1 GENE ENCODES FOR MANY DIFFERENT
CC       UDPGT ISOZYMES WHICH HAVE A DIFFERENT N-TERMINAL DOMAIN AND A
CC       COMMON C-TERMINAL DOMAIN OF 245 RESIDUES.
CC   -!- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D38069; BAA07263.1; -.
DR   EMBL; M34007; AAA42312.1; ALT_TERM.
DR   InterPro; IPR002213; -.
DR   Pfam; PF00201; UDPGT; 1.
DR   PROSITE; PS00375; UDPGT; 1.
KW   Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal;
KW   Multigene family; Microsome; Alternative splicing.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    531       UDP-GLUCURONOSYLTRANSFERASE 1-5.
FT   TRANSMEM    489    505       POTENTIAL.
FT   CARBOHYD    116    116       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    131    131       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    139    139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    431    431       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   531 AA;  59993 MW;  04148C1BA6CAAC80 CRC64;


  Query Match             5.4%;  Score 110.5;  DB 1;  Length 531;
  Best Local Similarity   20.4%;  Pred. No. 0.43;
  Matches   86;  Conservative   73;  Mismatches  126;  Indels  137;  Gaps   24;

Qy       4 QNKKILIITGSFGNGHMQVTQ-----SIVNQLNDMNLDHL-SVIEHDLFMEAHPILTSIC 57
           :| |::  ||::    :: ::     | |   : ||| |  |:::|
Db      92 RNIKVVFETGNYVKTFLETSEILKNISTVLLRSCMNLLHNGSLLQH-------------- 137

Qy      58 KKWYINSFKYFRNMYKGFYYSRPDKLDKC---FYKYYGLNKLINLLIKEKPDLILLTF-P 113
              :||          |     | :  |     || |:           | :  | : |
Db     138 ----LNS--------SSFDMVLTDPVIPCGQVLAKYLGI-----------PTVFFLRYIP 174

Qy     114 TPVMSVLTE--QFNINIP-VATVMTDYRLH----KNWITPYS---------TRYYVATKE 157
             : |  |:  : :  || : |:::|:       || : | :          |||     |
Db     175 CGIDSEATQCPKPSSYIPNLLTMLSDHMTFLQRVKNMLYPLALKYICHFSFTRYESLASE 234

Qy     158 TKQ-------------------DFI---DVGIDPSTVKVTGI------PIDNKFETPINQ 189
           |                     ||:     : |: | : ||      |:  :||  :|
Db     235 LLQREVSLVEVLSHASVWLFRGDFVFDYPRPVMPNMVFIGGINCVIKKPLSQEFEAYVN- 293

Qy     190 KQWLIDNNLDPDKQTILMSAGAFGV---SKGFDTMITDILAKSANAQVVMICGKSKELKR 246
                            ::|  |:   | |  :|:::|  | |  ::    |:   : :
Db     294 ------------------ASGEHGIVVFSLG--SMVSEIPEKKA-MEIAEALGR---IPQ 329

Qy     247 SLTAKFKLTRMYLILGYTKHMNEWMASSQLM--------ITKPGGITITEGFARCIPMIF 298
           :|  ::  ||   :   |  : :|:  : |:         ||  |   | ||    :||:
Db     330 TLLWRYTGTRPSNLAKNT-ILVKWLPQNDLLGHPKARAFITHSGSHGIYEGICNGVPMVM 388

Qy     299 LNPAPGQELENAFYFEEKGFG-------KIADTPEEAIKIVASLTNGNEQLTNMISTMEQ 351
           : |  | :::||   | :| |          ||  | |:| | :  :  | :   :|:: :
Db     389 M-PLFGDQMDNAKRMETRGAGVTLNVLEMTADDLENALKTVINNKSYKENIMR-LSSLHK 446

Qy     352 DK 353
           |:
Db     447 DR 448


Search completed: June 29, 2001, 09:04:11
Job time: 303 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 29, 2001, 09:05:02 ; Search time 38.9 Seconds
(without alignments)
1329.853 Million cell updates/sec

Title: US-09-668-788-4
Perfect score: 2048
Sequence: 1 MVTQNKKILIITGSFGNGHM.....SSQPQEIYGKVPLYARFFVK 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_16:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2048	100.0	391	2	O86492	O86492 staphylococ
2	309	15.1	464	10	Q9SI93	Q9si93 arabidopsis
3	309	15.1	465	10	Q9FZL5	Q9fzl5 arabidopsis
4	299	14.6	411	2	Q9RVF3	Q9rvf3 deinococcus
5	296.5	14.5	468	10	O82730	O82730 arabidopsis
6	284	13.9	533	8	Q9MU68	Q9mu68 arabidopsis
7	284	13.9	533	10	O81770	O81770 arabidopsis
8	283	13.8	525	10	P93115	P93115 cucumis sat
9	279.5	13.6	522	10	Q9SM44	Q9sm44 spinacia ol
10	279.5	13.6	530	10	Q9FZL4	Q9fzl4 glycine max
11	278	13.6	535	10	Q9FZL3	Q9fzl3 nicotiana t
12	245	12.0	373	2	O34625	O34625 bacillus su
13	230	11.2	374	2	Q9KBH0	Q9kbh0 bacillus ha
14	194.5	9.5	384	2	Q9EX00	Q9ex00 streptomyce
15	139	6.8	374	2	Q9S519	Q9s519 escherichia
16	124	6.1	363	2	Q9K9T0	Q9k9t0 bacillus ha
17	121	5.9	359	2	O87161	O87161 vibrio chol
18	116	5.7	530	6	O18777	O18777 oryctolagus
19	115.5	5.6	398	1	Q58652	Q58652 methanococc
20	113.5	5.5	765	3	Q01072	Q01072 kluyveromyc
21	112.5	5.5	383	2	O51410	O51410 borrelia bu
22	112	5.5	360	2	Q9FB02	Q9fb02 streptococc
23	109	5.3	396	2	O06038	O06038 lactococcus
24	108.5	5.3	363	1	Q9YCS0	Q9ycs0 aeropyrum p
25	108.5	5.3	495	5	O16917	O16917 caenorhabdi
26	106	5.2	533	5	Q23334	Q23334 caenorhabdi
27	106	5.2	1082	14	Q9WT34	Q9wt34 human herpe
28	105	5.1	372	2	Q9L4H3	Q9l4h3 brevibacter
29	104	5.1	530	11	P97886	P97886 rattus norv
30	104	5.1	1196	2	Q45916	Q45916 clostridium
31	104	5.1	1196	2	Q53550	Q53550 clostridium
32	104	5.1	1196	9	Q38197	Q38197 clostridium
33	104	5.1	1196	9	Q9ZX77	Q9zx77 clostridium
34	102.5	5.0	534	5	Q18629	Q18629 caenorhabdi
35	101.5	5.0	368	2	O66802	O66802 aquifex aeo
36	101.5	5.0	491	14	Q9DHG6	Q9dhg6 yaba-like d
37	101	4.9	530	5	Q93242	Q93242 caenorhabdi
38	101	4.9	987	3	Q9HFH5	Q9hfh5 pichia etch
39	101	4.9	1086	14	Q69054	Q69054 human herpe
40	100.5	4.9	238	2	Q9FDJ3	Q9fdj3 mycoplasma
41	100.5	4.9	588	2	O67183	O67183 aquifex aeo
42	100	4.9	502	10	Q9XG35	Q9xg35 guillardia
43	99.5	4.9	529	6	O18736	O18736 bos taurus
44	99.5	4.9	530	6	Q9TSM0	Q9tsm0 macaca fasc
45	99.5	4.9	533	6	O46423	O46423 felis silve

ALIGNMENTS

RESULT 1
O86492
ID O86492 PRELIMINARY; PRT; 391 AA.
AC O86492;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE YPFP PROTEIN.
GN YPFP.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=98313013; PubMed=9650993;
RA Ludovice A.M., Wu S., de Lencastre H.;
RT "Molecular cloning and DNA sequencing of the Staphylococcus aureus
RT UDP-N-acetylmuramyl tripeptide synthetase (murE) gene, essential for
RT the optimal expression of methicillin resistance.";
RL Microb. Drug Res. 4:85-90(1998).
DR EMBL; Y14370; CAA74741.1; -.
SQ SEQUENCE 391 AA; 44703 MW; 40CDDC37A2627C6C CRC64;

Query Match 100.0%; Score 2048; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.6e-151;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MVTQNKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKW 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MVTQNKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKW 60

Qy      61 YINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDLILLTFPTPVMSVL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 YINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDLILLTFPTPVMSVL 120

Qy     121 TEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGIPID 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGIPID 180
```

```
Qy    181 NKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSANAQVVMICGK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 NKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSANAQVVMICGK 240

Qy    241 SKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITITEGFARCIPMIFLN 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 SKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITITEGFARCIPMIFLN 300

Qy    301 PAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTNGNEQLTNMISTMEQDKIKYATQT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 PAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTNGNEQLTNMISTMEQDKIKYATQT 360

Qy    361 ICRDLLDLIGHSSQPQEIYGKVPLYARFFVK 391
          ||||||||||||||||||||||||||||||||
Db    361 ICRDLLDLIGHSSQPQEIYGKVPLYARFFVK 391
```

RESULT 2
Q9SI93
ID Q9SI93 PRELIMINARY; PRT; 464 AA.
AC Q9SI93;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE PUTATIVE MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE.
GN AT2G11810.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
DR EMBL; AC007187; AAD28678.1; -.
DR InterPro; IPR001296; -.
DR Pfam; PF00534; Glycos_transf_1; 1.
SQ SEQUENCE 464 AA; 52859 MW; FEC2B424CFBBA136 CRC64;

Query Match 15.1%; Score 309; DB 10; Length 464;
Best Local Similarity 25.4%; Pred. No. 5e-16;
Matches 104; Conservative 83; Mismatches 178; Indels 44; Gaps 16;

```
Qy      6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSF 65
          | :||:    | ||    ::| :       |   :|  |:: |          | :|
Db     72 KTVLILMSDTGGGHRASAEAIRDAFKIEFGDDYRIIIKDVWKE--------YTGWPLNDM 123

Qy     66 -KYFRNMYK-------GFYYSRPDKLDKCFYK----YYGLNKLINLLIKEKPDLILLTFP 113
           : :: | |        |: : |  : | :      ||   ::   |:: |||:|:   |
Db    124 ERQYKFMVKHVGLWSVAFHGTSPKWIHKSYLSALAAYYA-KEIEAGLMEYKPDIIISVHP 182

Qy    114 ----TPVMSVLTEQFNINIPVATVMTDYR-LHKNWITPYSTRYYVATKETKQDFIDVGID 168
              |:  :  :  :  :   ||:||    |: |     :| |  :|| :  :  |:|
Db    183 LMQHIPLWVMKWQGLHKKVIFVTVITDLNTCHRTWFHHGVSRCYCPSKEVAKRALVDGLD 242

Qy    169 PSTVKVTGIPIDNKF-ETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMIT--DI 225
          | ::| |:|:   |  | :|: :   :  :| :   :|:  |  |:     | :  |
Db    243 DSQIRVFGLPVRPSFPRTILNKNELRKELEIDLNLPAVLLMGGGEGMGPVQKTALALGDS 302

Qy    226 L--AKSAN--AQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLMITK 280
          |  :| :|   |:::|||::| |  :| : ::|:     : |:   | :|| :   :|||
Db    303 LYNSKESNPIGQLIVICGRNKVLASTLASHEWKIP--VKVRGFETQMEKWMGACDCIITK 360
```

```
Qy    281 PGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVAS-LTNGN 339
          |  || |     :|:|  :  |||| |  |  : | |    :|:|  ||||   :|
Db    361 AGPGTIAEALICGLPIILNDYIPGQEKGNVPYVVDNGAGVFTRSPKETAKIVADWFSNNK 420

Qy    340 EQLTNMISTMEQDKIKYATQTICRDLLDLIGHSSQPQEIYGKVPLYARF 388
          |:|      | :: :| :       |::  | | || |:   ::||:  |
Db    421 EEL----KKMSENALKLSQPEAVFDIVKDIHHLSQQQQ---RIPLFNEF 462
```

RESULT 3
Q9FZL5
ID Q9FZL5 PRELIMINARY; PRT; 465 AA.
AC Q9FZL5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE MGDG SYNTHASE TYPE C.
GN MGDC.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Awai K., Marechal E., Block M.A., Takamiya K., Joyard J., Ohta H.;
RT "The Multigenic Family of MGDG synthases.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB047398; BAB12041.1; -.
SQ SEQUENCE 465 AA; 52990 MW; 05E0157012E50A14 CRC64;

Query Match 15.1%; Score 309; DB 10; Length 465;
Best Local Similarity 25.4%; Pred. No. 5e-16;
Matches 104; Conservative 83; Mismatches 178; Indels 44; Gaps 16;

```
Qy      6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSF 65
          | :||:    | ||    ::| :       |   :|  |:: |          | :|
Db     73 KTVLILMSDTGGGHRASAEAIRDAFKIEFGDDYRIIIKDVWKE--------YTGWPLNDM 124

Qy     66 -KYFRNMYK-------GFYYSRPDKLDKCFYK----YYGLNKLINLLIKEKPDLILLTFP 113
           : :: | |        |: : |  : | :      ||   ::   |:: |||:|:   |
Db    125 ERQYKFMVKHVGLWSVAFHGTSPKWIHKSYLSALAAYYA-KEIEAGLMEYKPDIIISVHP 183

Qy    114 ----TPVMSVLTEQFNINIPVATVMTDYR-LHKNWITPYSTRYYVATKETKQDFIDVGID 168
              |:  :  :  :  :   ||:||    |: |     :| |  :|| :  :  |:|
Db    184 LMQHIPLWVMKWQGLHKKVIFVTVITDLNTCHRTWFHHGVSRCYCPSKEVAKRALVDGLD 243

Qy    169 PSTVKVTGIPIDNKF-ETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMIT--DI 225
          | ::| |:|:   |  | :|: :   :  :| :   :|:  |  |:     | :  |
Db    244 DSQIRVFGLPVRPSFPRTILNKNELRKELEIDLNLPAVLLMGGGEGMGPVQKTALALGDS 303

Qy    226 L--AKSAN--AQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLMITK 280
          |  :| :|   |:::|||::| |  :| : ::|:     : |:   | :|| :   :|||
Db    304 LYNSKESNPIGQLIVICGRNKVLASTLASHEWKIP--VKVRGFETQMEKWMGACDCIITK 361

Qy    281 PGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVAS-LTNGN 339
          |  || |     :|:|  :  |||| |  |  : | |    :|:|  ||||   :|
Db    362 AGPGTIAEALICGLPIILNDYIPGQEKGNVPYVVDNGAGVFTRSPKETAKIVADWFSNNK 421

Qy    340 EQLTNMISTMEQDKIKYATQTICRDLLDLIGHSSQPQEIYGKVPLYARF 388
          |:|      | :: :| :       |::  | | || |:   ::||:  |
Db    422 EEL----KKMSENALKLSQPEAVFDIVKDIHHLSQQQQ---RIPLFNEF 463
```

RESULT 4
Q9RVF3
ID Q9RVF3 PRELIMINARY; PRT; 411 AA.
AC Q9RVF3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

```
DE   CELL WALL SYNTHESIS PROTEIN, PUTATIVE.
GN   DR1076.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001958; AAF10649.1; -.
DR   TIGR; DR1076; -.
SQ   SEQUENCE   411 AA;  45017 MW;  D825DE52B8801437 CRC64;


  Query Match             14.6%;  Score 299;  DB 2;  Length 411;
  Best Local Similarity   25.3%;  Pred. No. 2.5e-15;
  Matches   84;  Conservative   63;  Mismatches  153;  Indels   32;  Gaps    8;

Qy       7 KILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSFK 66
           : | :: | | || |  |::     :  :: |   |||          |     |    :
Db      29 RALFMSVSLGAGHDQAQQAVKQAFAERGVELLGA-EHDSVEYLSTFERSFTVDLYEFELR 87

Qy      67 YFRNMYKGFYY----SRPDKLDKCFYKYYGLNKLINLLIKEKPDLILLTF--PTPVMSVL 120
             |  :|:|||:     :|  :    : : |:    : | : :|:::: :|  |  |    |
Db      88 YAPWLYRGFYWLTDQDQPWNIISRMFTWLGMGAFKDELRELRPEVVINSFWAPAAVCDTL 147

Qy     121 TEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGIPID 180
             |        ::|||| | :|    :    ||::||::  :: |: |  |:||||||
Db     148 RAQTGQRFLNCLIVTDYRAHLHWARRETDLLMVASEETRRQMLERGVRPEQVEVTGIPIS 207

Qy     181 NKFETPINQKQWLI------DNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSAN--- 231
             |   :   :| :      : :| |    :|:| |     :|      |:| :  |
Db     208 PAFREVLAADRWALRAELFSEMSLRPGVPLLLLSGG----GRGHYAAAADVLTELGNLGR 263

Qy     232 -AQVVMIC---GKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITIT 287
            ||::     |:  |     |           ||: : :   :|:| |:: | ||:|:
Db     264 AVQVLVPASRQGEGTETIGGATVHH--------LGFRRDLPRLLAASDLVVGKAGGLTVA 315

Qy     288 EGFARCIPMIFLNPAPGQELENAFYFEEKGFG 319
           |  |  :|::   | ||||  || : |  | |
Db     316 EATALGVPLVIYAPIPGQEEHNADFLERHGAG 347


RESULT   5
O82730
ID   O82730      PRELIMINARY;      PRT;   468 AA.
AC   O82730;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE (EC 2.4.1.46).
GN   MGD.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Awai K., Shimojima M., Masuda T., Takamiya K.I., Ohta H.;
RT   "cDNA cloning of an Arabidopsis monogalactosyldiacylglycerol synthase
RT   revealed that the enzyme belongs to an uridine diphosphate-utilizing
RT   glycosyltransferase superfamily.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ000331; CAA04005.1; -.
DR   Mendel; 33300; Arath;1856;33300.
DR   InterPro; IPR000566; -.
DR   PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW   Transferase; Glycosyltransferase.
SQ   SEQUENCE   468 AA;  52726 MW;  58F56C3C1C383C24 CRC64;


  Query Match             14.5%;  Score 296.5;  DB 10;  Length 468;
  Best Local Similarity   25.4%;  Pred. No. 4.7e-15;
  Matches  100;  Conservative   78;  Mismatches  173;  Indels   43;  Gaps   14;

Qy       4 QNKKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYIN 63
           : | :|||:    | ||    ::| :       |   ||  |:: |          | :|
Db      67 RTKNVLILMSDTGGGHRASAEAIRDAFKIEFGDKYRVIVKDVWKE--------YTGWPLN 118

Qy      64 ----SFKYF---RNMYK-GFYYSRPDKLDKCFY----KYYGLNKLINLLIKEKPDLILLT 111
               |:|:      ::|  |: : |  :  |:      ||   ::   |:: ||::|:
Db     119 DMERSYKFMVKHVQLWKVAFHSTSPKWIHSCYLAAIAAYYA-KEVEAGLMEYKPEIIISV 177

Qy     112 FP----TPVMSVLTEQFNINIPVATVMTDYR-LHKNWITPYSTRYYVATKETKQDFIDVG 166
             |      |:  :  ::    :   ||:||    |  |  |   | |  ::|  :  :  |
Db     178 HPLMQHIPLWVLKWQELQKRVLFVTVITDLNTCHPTWFHPGVNRCYCPSQEVAKRALFDG 237

Qy     167 IDPSTVKVTGIPIDNKFETPINQKQWL-IDNNLDPDKQTILMSAGAFGV------SKGFD 219
           :| | ||:| |:|:   |   :  |  |  :  :| | : :|:  |  |:        :|  :
Db     238 LDESQVRVFGLPVRPSFARAVLVKDDLRKELEMDQDLRAVLLMGGGEGMGPVKETAKALE 297

Qy     220 TMITDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLMI 278
             : |   :    |:|:|||::|:|  :| |  :|:      : |:   | :|| :  :|
Db     298 EFLYDKENRKPIGQMVVICGRNKKLASALEAIDWKIP--VKVRGFETQMEKWMGACDCII 355

Qy     279 TKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIV----AS 334
           || |  || |   | :|:|  :  ||||  |  |  | | |    :|:|  :||    ::
Db     356 TKAGPGTIAESLIRSLPIILNDYIPGQEKGNVPYVVENGAGVFTRSPKETARIVGEWFST 415

Qy     335 LTNGNEQLTNMISTMEQDKIKYATQTICRDLLDL 368
             |:  || ::    : | :   |  | :|: :|
Db     416 KTDELEQTSDNARKLAQPE---AVFDIVKDIDEL 446


RESULT   6
Q9MU68
ID   Q9MU68      PRELIMINARY;      PRT;   533 AA.
AC   Q9MU68;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE.
GN   MGD1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Jarvis P., Doermann P., Peto C.A., Lutes J., Benning C., Chory J.;
RT   "Galactolipid-Deficiency and Abnormal Chloroplast Development in the
RT   Arabidopsis MGD Synthase 1 Mutant.";
RL   Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2000).
DR   EMBL; AF241797; AAF65066.1; -.
KW   Chloroplast.
SQ   SEQUENCE   533 AA;  58537 MW;  E581E67317CB9CC8 CRC64;


  Query Match             13.9%;  Score 284;  DB 8;  Length 533;
  Best Local Similarity   25.5%;  Pred. No. 5.4e-14;
  Matches  103;  Conservative   66;  Mismatches  185;  Indels   50;  Gaps   13;
```

```
Qy      6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSF 65
          ||:||:    | ||    ::|    |    |   |   ||: :  |         |  |
Db    142 KKVLILMSDTGGGHRASAEAIRAAFNQEFGDEYQVFITDLWTDHTP--------WPFNQL 193

Qy     66 KYFRN-------MYKGFYYSRPDKL----DKCFYKYYGLNKLINLLIKEKPDLILLTFP- 113
               |       ::|  ||    ::    :       :   ::   |:| :||:|:   |
Db    194 PRSYNFLVKHGTLWKMTYYGTSPRIVHQSNFAATSTFIAREIAQGLMKYQPDIIISVHPL 253

Qy    114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
              |:  : ::     |   ||:||    |  |     || |  : |  :     |::
Db    254 MQHVPLRVLRSKGLLKKIVFTTVITDLSTCHPTWFHKLVTRCYCPSTEVAKRAQKAGLET 313

Qy    170 STVKVTGIPIDNKFETPINQKQWL-----IDNNLDPDKQTILMSAGAFGV------SKGF 218
          | :|| |:|:   |  |:  |  |     :| ||      :|:  |  |:      ::
Db    314 SQIKVYGLPVRPSFVKPVRPKVELRRELGMDENL----PAVLLMGGGEGMGPIEATARAL 369

Qy    219 DTMITDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLM 277
             : |     |  ||::|||::|:|:  |::  :|:     : |:   | | | :   :
Db    370 ADALYDKNLGEAVGQVLIICGRNKKLQSKLSSLDWKIP--VQVKGFITKMEECMGACDCI 427

Qy    278 ITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTN 337
          |||  |  || |   | :|:|     |||  |  |  | | || : :|:|  ||||
Db    428 ITKAGPGTIAEAMIRGLPIILNGYIAGQEAGNVPYVVVENGCGKFSKSPKEISKIVADWFG 487

Qy    338 GNEQLTNMISTMEQDKIKYATQ----TICRDLLDLI-GHSSQPQ 376
              :  :  | |: :: |       |  |: :|:   :| ||
Db    488 ---PASKELEIMSQNALRLAKPEAVFKIVHDMHELVRKKNSLPQ 528


RESULT  7
O81770
ID   O81770      PRELIMINARY;      PRT;   533 AA.
AC   O81770;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE - LIKE PROTEIN
DE   (MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE-LIKE PROTEIN) (MGDG SYNTHASE
DE   TYPE A).
GN   F28M20.30 OR AT4G31780 OR MGDA.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Hoheisel J.,
RA   Mewes H.W., Mayer K.F.X., Schueller C., Bevan M.;
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Rieger M., Mueller-Auer S., Zipp M., Schaefer M., Mewes H.W.,
RA   Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Terryn N., Ardiles W., Buysshaert C., Dasseville R., De Clerck R.,
RA   De Keyser A., Neyt P., Rouze P., Van Den Daele H., Villaroel R.,
RA   Gielen J., Van Montagu M., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Awai K., Marechal E., Block M.A., Takamiya K., Joyard J., Ohta H.;
RT   "The Multigenic Family of MGDG synthases.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031004; CAA19745.1; -.
DR   EMBL; AL161579; CAB79896.1; -.
DR   EMBL; AB047399; BAB12042.1; -.
DR   Mendel; 32401; Arath;1856;32401.
SQ   SEQUENCE   533 AA;  58537 MW;  E581E67317CB9CC8 CRC64;


  Query Match             13.9%;  Score 284;  DB 10;  Length 533;
  Best Local Similarity   25.5%;  Pred. No. 5.4e-14;
  Matches  103;  Conservative   66;  Mismatches  185;  Indels   50;  Gaps   13;

Qy      6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSF 65
          ||:||:    | ||    ::|    |    |   |   ||: :  |         |  |
Db    142 KKVLILMSDTGGGHRASAEAIRAAFNQEFGDEYQVFITDLWTDHTP--------WPFNQL 193

Qy     66 KYFRN-------MYKGFYYSRPDKL----DKCFYKYYGLNKLINLLIKEKPDLILLTFP- 113
               |       ::|  ||    ::    :       :   ::   |:| :||:|:   |
Db    194 PRSYNFLVKHGTLWKMTYYGTSPRIVHQSNFAATSTFIAREIAQGLMKYQPDIIISVHPL 253

Qy    114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
              |:  : ::     |   ||:||    |  |     || |  : |  :     |::
Db    254 MQHVPLRVLRSKGLLKKIVFTTVITDLSTCHPTWFHKLVTRCYCPSTEVAKRAQKAGLET 313

Qy    170 STVKVTGIPIDNKFETPINQKQWL-----IDNNLDPDKQTILMSAGAFGV------SKGF 218
          | :|| |:|:   |  |:  |  |     :| ||      :|:  |  |:      ::
Db    314 SQIKVYGLPVRPSFVKPVRPKVELRRELGMDENL----PAVLLMGGGEGMGPIEATARAL 369

Qy    219 DTMITDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLM 277
             : |     |  ||::|||::|:|:  |::  :|:     : |:   | | | :   :
Db    370 ADALYDKNLGEAVGQVLIICGRNKKLQSKLSSLDWKIP--VQVKGFITKMEECMGACDCI 427

Qy    278 ITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTN 337
          |||  |  || |   | :|:|     |||  |  |  | | || : :|:|  ||||
Db    428 ITKAGPGTIAEAMIRGLPIILNGYIAGQEAGNVPYVVVENGCGKFSKSPKEISKIVADWFG 487

Qy    338 GNEQLTNMISTMEQDKIKYATQ----TICRDLLDLI-GHSSQPQ 376
              :  :  | |: :: |       |  |: :|:   :| ||
Db    488 ---PASKELEIMSQNALRLAKPEAVFKIVHDMHELVRKKNSLPQ 528


RESULT  8
P93115
ID   P93115      PRELIMINARY;      PRT;   525 AA.
AC   P93115;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE PRECURSOR (EC 2.4.1.46)
DE   (1,2-DIACYLGLYCEROL 3-BETA-GALACTOSYLTRANSFERASE).
OS   Cucumis sativus (Cucumber).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Cucurbitales; Cucurbitaceae; Cucumis.
OX   NCBI_TaxID=3659;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=AONAGAJIBAI;
RX   MEDLINE=97144442; PubMed=8990209;
RA   Shimojima M., Ohta H., Iwamatsu A., Masuda T., Shioi Y.,
RA   Takamiya K.-I.;
RT   "Cloning of the gene for monogalactosyldiacylglycerol synthase and its
RT   evolutionary origin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:333-337(1997).
CC   -!- FUNCTION: CATALYZES THE FORMATION OF MONOGALACTOSYLDIACYLGLYCEROL
CC       (MGDG) WHICH IS A MAJOR STRUCTURAL LIPID OF THE CHLOROPLAST.
CC   -!- CATALYTIC ACTIVITY: UDP-GALACTOSE + 1,2-DIACYLGLYCEROL = UDP + 3-
CC       BETA-D-GALACTOSYL-1,2- DIACYLGLYCEROL.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: CHLOROPLAST.
CC   -!- SIMILARITY: TO E.COLI AND B.SUBTILIS UDP-N-ACETYLGLUCOSAMINE--N-
CC       ACETYLMURAMYL-(PENTAPEPTIDE) PYROPHOSPHORYL-UNDECAPRENOL N-
CC       ACETYLGLUCOSAMINE TRANSFERASE.
DR   EMBL; U62622; AAC49624.1; -.
DR   Mendel; 12656; Cucsa;1856;12656.
KW   Transferase; Glycosyltransferase; Transit peptide; Chloroplast.
```

```
FT   TRANSIT       1    103       CHLOROPLAST.
FT   CHAIN       104    525       MONOGALACTOSYLDIACYLGLYCEROL SYNTHASE.
SQ   SEQUENCE   525 AA;  57862 MW;  74FE586082EC48BA CRC64;


  Query Match             13.8%;  Score 283;  DB 10;  Length 525;
  Best Local Similarity   25.3%;  Pred. No. 6.3e-14;
  Matches  105;  Conservative   65;  Mismatches  185;  Indels   60;  Gaps   13;

Qy       6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSF 65
           |::||:    | ||    ::|    |:   ::  |  ||: :  |        |  |
Db     136 KRVLILMSDTGGGHRASAEAIKAAFNEEFGNNYQVFITDLWTDHTP--------WPFNQL 187

Qy      66 KYFRN-------MYKGFYYSRPDKL----DKCFYKYYGLNKLINLLIKEKPDLILLTFP- 113
               |       ::|  ||    |:    :      :   ::   |:| :||:|:   |
Db     188 PRSYNFLVKHGTLWKMTYYVTAPKVIHQSNFAATSTFIAREVAKGLMKYRPDIIISVHPL 247

Qy     114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
               |:  : ::     |   ||:||    |  |     || |  :  |  :   :  |: |
Db     248 MQHVPIRILRSKGLLNKIVFTTVVTDLSTCHPTWFHKLVTRCYCPSTEVAKRALTAGLQP 307

Qy     170 STVKVTGIPIDNKFETPIN-----QKQWLIDNNLDPDKQTILMSAGAFGV------SKGF 218
           | :|| |:|:   |  ||      :|:  :| ||      :|:  |  |:      :|
Db     308 SKLKVFGLPVRPSFVKPIRPKIELRKELGMDENL----PAVLLMGGGEGMGPIEATAKAL 363

Qy     219 DTMITDILAKSANAQVVMICGKSKELK---RSLTAKFKLTRMYLILGYTKHMNEWMASSQ 275
             : |        ||::||| :|:|    ||:  |  :     : |:   | | | :
Db     364 SKALYDENHGEPIGQVLVICGHNKKLAGRLRSIDWKVPVQ----VKGFVTKMEECMGACD 419

Qy     276 LMITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASL 335
             :||| |  || |   | :|:|  :   |||  |  |  | | || : :|:|   |||
Db     420 CIITKAGPGTIAEAMIRGLPIILNDYIAGQEAGNVPYVVENGCGKFSKSPKEIANIVAKW 479

Qy     336 TNGNEQLTNMISTMEQDKIKYATQ----TICRDLLDLIGHSSQPQEIYGKVPLYA 386
                 : :  | |: :: |       |  || :|:   |         || |:
Db     480 FGPK---ADELLIMSQNALRLARPDAVFKIVHDLHELVKQRS-------FVPQYS 524


RESULT   9
Q9SM44
ID   Q9SM44      PRELIMINARY;      PRT;   522 AA.
AC   Q9SM44;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MGDG SYNTHASE A PRECURSOR (EC 2.4.1.46).
GN   MGD A.
OS   Spinacia oleracea (Spinach).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
OC   Caryophyllales; Chenopodiaceae; Spinacia.
OX   NCBI_TaxID=3562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99449603; PubMed=10518794;
RA   Miege C., Marechal E., Shimojima M., Awai K., Block M.A., Ohta H.,
RA   Takamiya K.I., Douce R., Joyard J.;
RT   "Biochemical and topological properties of type A MGDG synthase, a
RT   spinach chloroplast envelope enzyme catalyzing the synthesis of both
RT   prokaryotic and eukaryotic MGDG.";
RL   Eur. J. Biochem. 265:990-1001(1999).
DR   EMBL; AJ249607; CAB56218.1; -.
KW   Transit peptide; Transferase; Glycosyltransferase.
FT   TRANSIT       1     98
FT   CHAIN        99    522       MGDG SYNTHASE A.
SQ   SEQUENCE   522 AA;  57511 MW;  02E2B929732551A7 CRC64;


  Query Match             13.6%;  Score 279.5;  DB 10;  Length 522;
  Best Local Similarity   24.5%;  Pred. No. 1.2e-13;
  Matches   96;  Conservative   64;  Mismatches  191;  Indels   41;  Gaps   11;

Qy       6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSF 65
           |::||:    | ||    ::|    |:   |  |  ||: :  |        |  |
Db     131 KRVLILMSDTGGGHRASAEAIKAAFNEEFGDDYQVFVTDLWSEHTP--------WPFNQL 182

Qy      66 KYFRN-------MYKGFYYSRPDKL----DKCFYKYYGLNKLINLLIKEKPDLILLTFP- 113
               |       ::|  ||    ::    :      :   ::   |:| :||:|:   |
Db     183 PRSYNFLVKHGPLWKMMYYGTSPRVIHQSNFAATSVFIAREVARGLMKYQPDIIISVHPL 242

Qy     114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
               |:  :        |   ||:||    |  |     || |  : |  :     |: |
Db     243 MQHVPLRILRGRGLLEKIVFTTVVTDLSTCHPTWFHKLVTRCYCPSNEVAKRATKAGLQP 302

Qy     170 STVKVTGIPIDNKFETPINQKQWL-IDNNLDPDKQTILMSAGAFGV------SKGFDTMI 222
           | :|| |:|:   |  :  |  |  :  :|     :|:  |  |:      ::      :
Db     303 SQIKVYGLPVRPSFVRSVRPKNELRKELGMDEHLPAVLLMGGGEGMGPIEATARALGNAL 362

Qy     223 TDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLMITKP 281
           |        |:::|||::|:|   |::  :|:     : |:   : | | :   :|||
Db     363 YDANLGEPTGQLLVICGRNKKLAGKLSSIDWKIP--VQVKGFVTKIEECMGACDCIITKA 420

Qy     282 GGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTNGNEQ 341
           |  || |   | :|:|  :   |||  |  |  | | ||   :|:|  | |:
Db     421 GPGTIAEAMIRGLPIILNDYIAGQEAGNVPYVIENGIGKYLKSPKEIAKTVSQWFGPK-- 478

Qy     342 LTNMISTMEQDKIKYATQ----TICRDLLDLI 369
             | :  | |: :|:|        |  || :|:
Db     479 -ANELQIMSQNALKHARPDAVFKIVHDLDELV 509


RESULT  10
Q9FZL4
ID   Q9FZL4      PRELIMINARY;      PRT;   530 AA.
AC   Q9FZL4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MGDG SYNTHASE TYPE A.
GN   GMMGD A.
OS   Glycine max (Soybean).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC   Fabales; Fabaceae; Papilionoideae; Glycine.
OX   NCBI_TaxID=3847;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Awai K., Takamiya K., Ohta H.;
RT   "cDNA cloning of MGDG synthase from tobacco and soybean.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB047475; BAB11979.1; -.
SQ   SEQUENCE   530 AA;  57839 MW;  D33C37FD53E90218 CRC64;


  Query Match             13.6%;  Score 279.5;  DB 10;  Length 530;
  Best Local Similarity   25.0%;  Pred. No. 1.2e-13;
  Matches   99;  Conservative   69;  Mismatches  187;  Indels   41;  Gaps   13;

Qy       6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYIN-- 63
           ||:||:    | ||    ::|          |   |   ||: :  |        |  |
Db     139 KKVLILMSDTGGGHRASAEAIKAAFYQEFGDDYQVFVTDLWADHTP--------WPFNQL 190

Qy      64 --SFKYFRN---MYKGFYY-SRPDKLDKCFYKYYG---LNKLINLLIKEKPDLILLTFP- 113
             |: :       ::|  || : |  : :  :     |      ::   |:| :||:|:   |
Db     191 PRSYSFLVKHGPLWKMTYYGTAPRVVHQSNFAATGTFIAREVAKGLMKYQPDIIISVHPL 250

Qy     114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
               |:  : ::     |   ||:||    |  |     || |  : |  :  |  :  |:
Db     251 MQHVPLRILRSKGLLKKIVFTTVITDLSTCHPTWFHKLVTRCYCPTTDVAQRALKAGLQQ 310

Qy     170 STVKVTGIPIDNKFETPINQKQWL-IDNNLDPDKQTILMSAGAFGV------SKGFDTMI 222
           | :|: |:|:   |  |:  |  |  :  :| |   :|:  |  |:      ::      :
Db     311 SQIKIFGLPVRPSFVKPVQPKDELRRELGMDEDLPAVLLMGGGEGMGPIEATARALGDSL 370

Qy     223 TDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLMITKP 281
             |    :   |:::|||::|:|   |::  :|:     : |:   | | | :   :|||
Db     371 YDANLGEPTGQLLVICGRNKKLAGKLSSIDWKIP--VQVKGFVTKIEECMGACDCIITKA 428
```

```
Db    371 YDENIGAPVGQILVICGRNKKLANKLSSINWKVP--VQVKGFVTKMEECMGACDCIITKA 428

Qy    282 GGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTNGNEQ 341
          |  || |   | :|:|  :   ||| |  |  | | || : :|::  ||||
Db    429 GPGTIAEAQIRGLPIILNDYIAGQEAGNVPYVVENGCGKFSKSPKDIAKIVAEWFGPK-- 486

Qy    342 LTNMISTMEQDKIKYATQ----TICRDLLDLIGHSS 373
            :  | |: :: |       |  || :|:   |
Db    487 -AYELQQMSQNALRLARPDAVFKIVHDLHELVRQRS 521


RESULT  11
Q9FZL3
ID   Q9FZL3      PRELIMINARY;      PRT;   535 AA.
AC   Q9FZL3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MGDG SYNTHASE TYPE A.
GN   NTMGD A.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC   Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Awai K., Takamiya K., Ohta H.;
RT   "cDNA cloning of MGDG synthase from tobacco and soybean.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB047476; BAB11980.1; -.
SQ   SEQUENCE   535 AA;  59589 MW;  B1B2067E86EDE477 CRC64;

  Query Match             13.6%;  Score 278;  DB 10;  Length 535;
  Best Local Similarity   24.5%;  Pred. No. 1.6e-13;
  Matches   98;  Conservative   78;  Mismatches  182;  Indels   42;  Gaps   13;

Qy      6 KKILIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYIN-- 63
          ||:||:    | ||    ::| :  |:   |   |   ||: |  |         |  |
Db    145 KKVLILMSDTGGGHRASAEAIRSAFNEEFGDKYQVFITDLWTEHTP--------WPFNQL 196

Qy     64 --SFKYF---RNMYKGFYYSRPDKL----DKCFYKYYGLNKLINLLIKEKPDLILLTFP- 113
            |: :     :::: ||:   :|    :       :   ::   |:| :||:|:   |
Db    197 PRSYNFLVKHGSLWRMTYYATAPRLVHQTNFAATSTFIAREVAKGLMKYQPDIIISVHPL 256

Qy    114 ---TPVMSVLTEQFNINIPVATVMTDY-RLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
              |:  : ::     |   ||:||    |  |     || |  ::|  :  :  |: |
Db    257 MQHVPLRILRSKGLLKKIIFTTVITDLSTCHPTWFHKLVTRCYCPSEEVAKRALRAGLKP 316

Qy    170 STVKVTGIPIDNKFETPINQKQWL-IDNNLDPDKQTILMSAGAFGV------SKGFDTMI 222
          :|| |:|:   |  |:  |  |  :  ::     :|:  |  |:       ::     :
Db    317 YQLKVYGLPVRPSFVKPVPPKVELRKELGMEEHLPAVLLMGGGEGMGPIEATARALGDAL 376

Qy    223 TDILAKSANAQVVMICGKSKELKRSLTA-KFKLTRMYLILGYTKHMNEWMASSQLMITKP 281
          | :      ||::|||::|:|   ||: ::|:     : |:   | | | :   :|||
Db    377 YDEIHGEPIGQVLVICGRNKKLFNRLTSVQWKIP--VQVKGFVTKMEECMGACDCIITKA 434

Qy    282 GGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTNGNEQ 341
          |  || |   | :|:|  :   ||| |  |  | | || : :|::   |||       :
Db    435 GPGTIAEAVIRGLPIILNDYIAGQEAGNVPYVIENGCGKFSKSPKKIANIVAQWFGPRQ- 493

Qy    342 LTNMISTMEQDKIKYATQ----TICRDLLDLIGHSS-QPQ 376
            : :  | |: :: |       |  |: :|:   : ||
Db    494 --DELRIMSQNALRLARPDAVFKIVHDMHELVRQRNFEPQ 531


RESULT  12
O34625
ID   O34625      PRELIMINARY;      PRT;   373 AA.
AC   O34625;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   YKON.
GN   YKON.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Devine K.M.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J.; Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ002571; CAA05612.1; -.
DR   EMBL; Z99111; CAB13208.1; -.
DR   EMBL; Z99110; CAB13192.1; -.
SQ   SEQUENCE   373 AA;  43480 MW;  62FFCF2BDC598CDC CRC64;

  Query Match             12.0%;  Score 245;  DB 2;  Length 373;
  Best Local Similarity   23.5%;  Pred. No. 3.5e-11;
  Matches   88;  Conservative   75;  Mismatches  168;  Indels   44;  Gaps   14;

Qy      6 KKILIITG-SFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINS 64
          | |||     |   ||  |  ::  :|       |:  : |:|  ::  |  :    |:
Db      2 KNILIFPFLSISTGHHHVADALQAELESQG---LAAEKIDIFSHSYRRLEKLSSVAYLKW 58

Qy     65 FKYFRNMYKGFYYSRPDKL--------DKCFYKYYGL--NKLINLLIKEKPDLILLT--F 112
             :||   | | |     :|        || :: |  :   :: ::| :::||:   |
Db     59 IQYFPKTYSGIY-----RLLACGEFQHDKRYFMYECVFTQQMRHILQEKQPDIAFCTHAL 113

Qy    113 PTPVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTV 172
          |: ::: |  ::  |: |  | ||: ::: |       ::| : | |:  |  ||| : :
Db    114 PSYLLNRLKPEYP-NLTVVNVYTDFFVNQLWGRKNIDYHFVPSTEVKKQLISEGIDQNNI 172
```

```
Qy    173 KVTGIPIDNKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSANA 232
          :||||:   ||          |         ||::: |: ||  |    : : |:
Db    173 YLTGIPVHQNFEME------SADTLQHHPPYTIIITGGSMGVG-GILKWVQE-LSPGGKI 224

Qy    233 QVVMICGKSKELKRSL-TAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITITEGFA 291
          ::||::::|   : :    |      |     ||     :  ::|||||:||:|
Db    225 LYKILCGRNEKLYSYVKSLHHPLIEAIPYLHSKAEMNRLYEQATGIMTKPGGVTISECLQ 284

Qy    292 RCIPMIFLNPAPGQELENAFYFEEKGFGKIAD----TPEEAIKIVASLTNGNEQL----- 342
          : :|:   :  ||||  |     |:    : |      ::| : :|:    |||:
Db    285 KRLPVFIYHALPGQEEMNLNLLHERKL--VTDMRNWDMQKAEEYIAAFFQSNEQMKEYKK 342

Qy    343 --TNMISTMEQDKIK 355
              :  |   |||
Db    343 HVNGYLGEMSDRKIK 357


RESULT  13
Q9KBH0
ID   Q9KBH0      PRELIMINARY;      PRT;   374 AA.
AC   Q9KBH0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BH1957 PROTEIN.
GN   BH1957.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RA   Takami H., Nakasone K., Takaki Y.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP001513; BAB05676.1; -.
SQ   SEQUENCE   374 AA;  42361 MW;  DA96A0EA33C5AC65 CRC64;

  Query Match           11.2%;  Score 230;  DB 2;  Length 374;
  Best Local Similarity  21.8%;  Pred. No. 5.1e-10;
  Matches   83;  Conservative   82;  Mismatches  165;  Indels   50;  Gaps   14;

Qy      9 LIITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSFKYF 68
          || : | |||| |  :::  |:   |  :    |  | |     | ||          |:|  |
Db      6 LIFSASIGNGHNQAAKAL--QVEFQNKGYQPEI-IDTFYSLSPALHKFMLTSYVNLLKVG 62

Qy     69 RNMYKGFYYSRPDK------LDKCFYKYYGLNKLINLLIKEKPDLILLTFP--TPVMSVL 120
            :::  |: : :|      ||:  :  : :  |   :   :  :: | |  |  :  |
Db     63 PRIWQKIYF-QAEKYPLFLFLDQ--FATFFVESLHATVKSNRCSFLVSTHPFVTAFLVRL 119

Qy    121 TEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDV-GIDPSTVKVTGIPI 179
            :  :|:|: ||:||: ||  :: |    |: :      ||  :  :       |||||
Db    120 KSKKQLNLPLYTVITDFVLHPAYLRPEIDGYFTSDPNF-TDFAKLNNVSDDRFFPTGIPI 178

Qy    180 DNKFETPINQKQWLIDNN--LDPDKQTILMSAGAFGVSKGFDTMITDILAKSANAQVVMI 237
          |     |:| :| : |:  ||| :: :::: |  |::  :  :|  :       |:: :
Db    179 PNL--ESIDQPKWKVRNDLGLDPQRKVLIIAGGGIGLT-NYAQVIRALECLPEPIQLLCM 235

Qy    238 CGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITITEGFARCIPMI 297
          | : ::|  : :: |      :: :|   :::| ||:|: |      |:|
Db    236 IGHNYQVKEKI-SRIKSKHELKVIEFTDKFLLYLKASDAILSKAGGLTMAESLVCETPII 294

Qy    298 FLNPAPGQELENAFY----------------------FEEKGFGKIADTPEEAIKIVA 333
            | || |  || :                        :||  || : :   :  |  |
Db    295 IHQPVPGHEEHNAKFLIDAGAALRVKGSKEIPTTIKRVLYEEACFGPMIENARKLKKPNA 354

Qy    334 SLTNGNEQLTNMISTMEQDK 353
          :    || :  |:  ::: :
Db    355 A----NEIVEQMLLLVKEQQ 370


RESULT  14
Q9EX00
ID   Q9EX00      PRELIMINARY;      PRT;   384 AA.
AC   Q9EX00;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PUTATIVE SECRETED PROTEIN.
GN   2SCG38.19C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Saunders D.C., Harris D.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL445503; CAC13078.1; -.
SQ   SEQUENCE   384 AA;  40689 MW;  E64E01E05B1DDE40 CRC64;

  Query Match           9.5%;  Score 194.5;  DB 2;  Length 384;
  Best Local Similarity  20.2%;  Pred. No. 3.1e-07;
  Matches   72;  Conservative   73;  Mismatches  160;  Indels   51;  Gaps   11;

Qy      6 KKILIITGSFGNGHMQVTQSIVNQLND-----MNLDHLSVIEHDLFMEAHPILTSICKKW 60
          :::|:|: | | ||  |   :| :  :        :| |::: :          | :: : :
Db     15 RRLLVISASMGAGHDTVAAELVRRARERGDTAQTVDVLALLPYG--------LGAVLRCF 66

Qy     61 YINSFKYFRNMYKGFY--YSRPDKLDKCFYKYYGL--------------NKLINLLIKEK 104
          |  | :|   |   |  : ||          |                  ::|  |  :
Db     67 YRGSVRHFPWAYAALYRLFLRP----------GAGRRPSGTPLARLAGDRLRELAARTG 115

Qy    105 PDLILLTF--PTPVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDF 162
           |:::  |      :   | ::  : :|   :: |: ||: |: | :       |:|  ::
Db    116 ADVVVPVFHLGAQLTGHLRDRGLLPVPSVVLVIDFELHRQWLHPGNDHCLCLTEEAAREA 175

Qy    163 I-DVGIDPSTVKVTGIPIDNKFETPINQKQWL-IDNNLDPDKQTILMSAGAFGVSKGFDT 220
          : |   |      |  :   |    ||    : | | :  :::||||:||    |
Db    176 RGNTGTPAETCGPVVAPEFSAGRVP-GAAQWRETFDRLGPGRPAVVLSAGAWGVGSHLDA 234

Qy    221 MITDILAKSANAQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITK 280
          :  |        |::|| :: |:|:|:      |   | ||:  |   : ::: :|
Db    235 TVR--LLVDHGYLPVVLCGDNQRLRRTLSG----TPGVLALGWVTDMPGLLHAARALIDN 288

Qy    281 PGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLT 336
          | |  :  |  :|::   | ||   :        |  ::|:     :: :| ||
Db    289 AAGQTAVQALAAGLPVVGHRPIPGHGADGVRRMAALGVSEVAEDRTALLEALARLT 344


RESULT  15
Q9S519
ID   Q9S519      PRELIMINARY;      PRT;   374 AA.
AC   Q9S519;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PUTATIVE GLYCOSYL TRANSFERASE.
```

```
GN   WBDM.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=STOKE W, M92;
RX   MEDLINE=96060831; PubMed=7590310;
RA   Bastin D.A., Reeves P.R.;
RT   "Sequence and analysis of the O antigen gene (rfb) cluster of
RT   Escherichia coli O111.";
RL   Gene 164:17-23(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=STOKE W, M92;
RX   MEDLINE=98449835; PubMed=9774562;
RA   Wang L., Curd H., Qu W., Reeves P.R.;
RT   "Sequencing of Escherichia coli O111 O-antigen gene cluster and
RT   identification of O111-specific genes.";
RL   J. Clin. Microbiol. 36:3182-3187(1998).
DR   EMBL; AF078736; AAD46732.1; -.
DR   InterPro; IPR001296; -.
DR   Pfam; PF00534; Glycos_transf_1; 1.
KW   Transferase.
SQ   SEQUENCE   374 AA;  42192 MW;  54C0C1EFBBF76B53 CRC64;


  Query Match             6.8%;  Score 139;  DB 2;  Length 374;
  Best Local Similarity   20.9%;  Pred. No. 0.0062;
  Matches   84;  Conservative   69;  Mismatches  163;  Indels   86;  Gaps   18;

Qy       6 KKILIITGSFGNGHMQVTQSIVNQL----NDMNLDHLSVIEHDLFMEAHPILTSICKKWY 61
           | : ||||    |   :   : :|:    :|:|:  |: |      |  |  |      |
Db       2 KIVYIITGLTCGGAEHLMTQLADQMFIRGHDVNIICLTGIS-----EVKP--TQNINIHY 54

Qy      62 INSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDLI----------L 109
           :|  | ||: ::  :                    ::  :::  |||:|
Db      55 VNMDKNFRSFFRALF----------------QVKKIIVALKPDIIHSHMFHANIFSRF 96

Qy     110 LTFPTPVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDP 169
           :    | : ::    | |      |  |||  :::   :|     :||  |:||     |
Db      97 IRMLIPAVPLICTAHNKNEGGNARMFCYRL-SDFLASITTN---VSKEAVQEFIARKATP 152

Qy     170 STVKVTGIP---IDNKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITD-- 224
              |:  ||     |||:  || ::   |     |   :|:: |    :| :  ::
Db     153 KN-KIVEIPNFINTNKFDFDINVRKKTRDAFNLKDSTAVLLAVGRLVEAKDYPNLLNAIN 211

Qy     225 --ILAKSANAQ--VVMICG----KSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQL 276
             ||:|::|    :::| |    ::| |   |  :  |      ||    : | | :: |
Db     212 HLILSKTSNCNDFILLIAGDGALRNKLL--DLVCQLNLVDKVFFLGQRSDIKELMCAADL 269

Qy     277 MITKPG----GITITEGFARCIPMIFLNPAPGQE----------LENAFYFEEKGFGKIA 322
             :        |: : |  |   |::  :    :|          : |     |    |||
Db     270 FVLSSEWEGFGLVVAEAMACERPVVATDSGGVKEVVGPHNDVIPVSNHILLAE----KIA 325

Qy     323 DT---PEEAIKIVASLTN----GNEQLTNMISTMEQDKIKYA 357
           :|    : | ||:          | :  ::|  |:   ||:
Db     326 ETLKIDDNARKIIGMKNREYIVSNFSIKTIVSEWERLYFKYS 367


Search completed: June 29, 2001, 09:05:04
Job time: 321 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 29, 2001, 08:59:39 ; Search time 24.11 Seconds
(without alignments)
326.692 Million cell updates/sec

Title: US-09-668-788-4
Perfect score: 2048
Sequence: 1 MVTQNKKILIITGSFGNGHM.....SSQPQEIYGKVPLYARFFVK 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	119	5.8	352	2	US-08-751-474-2	Sequence 2, Appli
2	111.5	5.4	534	5	PCT-US92-00282-4	Sequence 4, Appli
3	109.5	5.3	531	5	PCT-US92-00282-6	Sequence 6, Appli
4	104.5	5.1	529	5	PCT-US92-00282-7	Sequence 7, Appli
5	102.5	5.0	934	1	US-08-457-176-2	Sequence 2, Appli
6	102.5	5.0	934	1	US-08-457-175-2	Sequence 2, Appli
7	102.5	5.0	934	3	US-08-709-784-1	Sequence 1, Appli
8	101.5	5.0	376	2	US-08-846-762-85	Sequence 85, Appl
9	99.5	4.9	533	5	PCT-US92-00282-3	Sequence 3, Appli
10	98.5	4.8	2366	1	US-08-480-604A-10	Sequence 10, Appl
11	98.5	4.8	2366	2	US-08-405-496A-10	Sequence 10, Appl
12	93	4.5	531	5	PCT-US92-00282-5	Sequence 5, Appli
13	92	4.5	648	1	US-08-451-715A-4	Sequence 4, Appli
14	91	4.4	515	4	US-08-942-012B-32	Sequence 32, Appl
15	89	4.3	937	4	US-09-005-180A-4	Sequence 4, Appli
16	86	4.2	2958	4	US-08-894-344C-2	Sequence 2, Appli
17	85	4.2	2496	4	US-09-125-028-2	Sequence 2, Appli
18	84	4.1	514	1	US-07-872-644-17	Sequence 17, Appl
19	84	4.1	514	1	US-08-297-494-17	Sequence 17, Appl
20	84	4.1	514	1	US-08-297-510-17	Sequence 17, Appl
21	84	4.1	514	1	US-08-479-532-17	Sequence 17, Appl
22	84	4.1	514	1	US-08-455-526-17	Sequence 17, Appl
23	84	4.1	514	1	US-08-455-525-17	Sequence 17, Appl
24	84	4.1	514	3	US-09-139-491-17	Sequence 17, Appl
25	84	4.1	514	5	PCT-US92-03222-17	Sequence 17, Appl
26	84	4.1	530	1	US-07-872-644-6	Sequence 6, Appli
27	84	4.1	530	1	US-08-297-494-6	Sequence 6, Appli
28	84	4.1	530	1	US-08-297-510-6	Sequence 6, Appli
29	84	4.1	530	1	US-08-479-532-6	Sequence 6, Appli
30	84	4.1	530	1	US-08-455-526-6	Sequence 6, Appli
31	84	4.1	530	1	US-08-455-525-6	Sequence 6, Appli
32	84	4.1	530	3	US-09-139-491-6	Sequence 6, Appli
33	84	4.1	530	5	PCT-US92-03222-6	Sequence 6, Appli
34	83	4.1	344	2	US-08-846-762-90	Sequence 90, Appl
35	83	4.1	753	2	US-08-867-941-20	Sequence 20, Appl
36	83	4.1	753	4	US-09-074-658-20	Sequence 20, Appl
37	82.5	4.0	506	4	US-08-942-012B-25	Sequence 25, Appl
38	82.5	4.0	506	4	US-08-942-012B-26	Sequence 26, Appl
39	82.5	4.0	506	6	5180581-2	Patent No. 5180581
40	82.5	4.0	872	1	US-08-766-014-2	Sequence 2, Appli
41	82.5	4.0	1205	1	US-07-908-245-2	Sequence 2, Appli
42	82	4.0	2165	1	US-08-514-975B-2	Sequence 2, Appli
43	82	4.0	2165	5	PCT-US95-12507-2	Sequence 2, Appli
44	81.5	4.0	625	2	US-08-532-547-9	Sequence 9, Appli
45	81.5	4.0	625	4	US-09-019-809-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-08-751-474-2
; Sequence 2, Application US/08751474
; Patent No. 5821335
; GENERAL INFORMATION:
; APPLICANT: Hoskins, JoAnn
; APPLICANT: Skatrud, Paul L.
; TITLE OF INVENTION: Biosynthetic Gene murG From
; TITLE OF INVENTION: Streptococcus Pneumoniae
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,474
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-9902
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 352 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-751-474-2

Query Match 5.8%; Score 119; DB 2; Length 352;
Best Local Similarity 20.4%; Pred. No. 0.00072;
Matches 79; Conservative 80; Mismatches 155; Indels 74; Gaps 19;

```
Qy       7 KILIITGSFGNGHMQVTQSI----------VNQLND-MNLDHLSVIEHDLFMEAHPILTS 55
           | :: ||    ||: :   :          |: : |   ::|  :::  | :  | | |
Db       2 KKIVFTGGGTVGHVTLNLLLMPKFIEDGWEVHYIGDKRGIEHQEILKSGLDVTFHSIATG 61

Qy      56 ICKKWYINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLIKEKPDLI-----LL 110
```

```
Db     62 KLRR----------------YFSWQNMLD-VFKVGWGIVQSLFIMLRLRPQTLFSKGGFV 104

Qy    111 TFPTPVMSVLTEQFNINIPVATVMTDYR--LHKNWITPYSTRYYVATKETKQDFIDVGID 168

Db    105 SVP-PVIAARVS----GVPVFIHESDLSMGLANKIAYKFATKMY-STFEQASSLSKVEHV 158

Qy    169 PSTVKVTGIPIDNKFETPINQKQWLID--NNLDPDKQTILMSAGAFGVSKGFDTMITDIL 226

Db    159 GAVTKVS----DQKNPEP----DELVDIQTHFNHKLPTVLFVGGSAG-ARVFNQLVTDHK 209

Qy    227 AK-SANAQVVMICGKS--KELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGG 283

Db    210 KELTERYNIINLTGDSSLNELSQNLFRVDYVTDLYQPL---------MELADVVVTRGGA 260

Qy    284 ITITEGFA----RCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTNGN 339

Db    261 NTIFELLAIAKLHVIVPLGREASRGDQIENAAYFVKKGYAE--DLQESDL----TLDSLE 314

Qy    340 EQLTNMISTMEQDKIKYATQTICRDLLD 367

Db    315 EKLSHLLSHKEDYQAKMKASKELKSLAD 342


RESULT   2
PCT-US92-00282-4
; Sequence 4, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  534 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US92-00282-4


  Query Match             5.4%;  Score 111.5;  DB 5;  Length 534;
  Best Local Similarity   20.8%;  Pred. No. 0.0079;
  Matches   60;  Conservative   54;  Mismatches  103;  Indels   71;  Gaps   12;

Qy     93 LNKLINLLIKEKPDLILLTFPTPVMSVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYY 152

Db    208 LQRVKNMLYPLALSYICHTFSAPYASLASELFQREVSVVDLVS----------------Y 251

Qy    153 VATKETKQDFI---DVGIDPSTVKVTGI------PIDNKFETPINQKQWLIDNNLDPDKQ 203

Db    252 ASVWLFRGDFVMDYPRPIMPNMVFIGGINCANGKPLSQEFEAYIN---------------- 296

Qy    204 TILMSAGAFGV---SKGFDTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLTRMYLI 260

Db    297 ----ASGEHGIVVFSLG--SMVSEIPEKKAMA----IADALGKIPQTVLWRYTGTRPSNL 346

Qy    261 LGYTKHMNEWMASSQLM--------ITKPGGITITEGFARCIPMIFLNPAPGQELENAFY 312

Db    347 ANNT-ILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMM-PLFGDQMDNAKR 404

Qy    313 FEEKGFG-------KIADTPEEAIKIVASLTNGNEQLTNMISTMEQDK 353

Db    405 METKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMR-LSSLHKDR 451


RESULT   3
PCT-US92-00282-6
; Sequence 6, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  531 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US92-00282-6


  Query Match             5.3%;  Score 109.5;  DB 5;  Length 531;
  Best Local Similarity   20.6%;  Pred. No. 0.012;
  Matches   78;  Conservative   69;  Mismatches  122;  Indels  109;  Gaps   21;

Qy     24 QSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSFKYFRNMYKGFYYSRPDKL 83

Db    132 KTLIQQLNSSSFD--VVLTDPVF----PCGALLAKYLQIPAVFFLRSVPCGIDY----EA 181

Qy     84 DKCFYKYYGLNKLINLLIKEKPDLI---LLTFPTPVMSVLTEQFNINIPVATVMTDYRLH 140

Db    182 TQC----------------PKPSSYIPNLLTMLSDHMTFLQRVKNMLYPLT---LKYICH 222
```

```
Qy    141 KNWITPYSTRYYVATK--ETKQDFIDV------------------GIDPSTVKVTGI-- 177
          : |||| :   :|::  : :   ::|                   | |: | : ||
Db    223 LS-ITPYES---LASELLQREMSLVEVLSHASVWLFRGDFVFDYPRPIMPNMVFIGGINC 278

Qy    178 ----PIDNKFETPINQKQWLIDNNLDPDKQTILMSAGAFGV---SKGFDTMITDILAKSA 230
              |:  :||  :|                  ::|  |:   | |  :|:::|  | |
Db    279 VIKKPLSQEFEAYVN------------------ASGEHGIVVFSLG--SMVSEIPEKKA 317

Qy    231 NAQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLM--------ITKPG 282
            ::    |:   : ::|  ::  ||   :   |  : :|:  : |:         ||  |
Db    318 -MEIAEALGR---IPQTLLWRYTGTRPSNLAKNT-ILVKWLPQNDLLGHPKARAFITHSG 372

Qy    283 GITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFG-------KIADTPEEAIKIVASL 335
            | ||    :||: : |  | :::||   | :| |          ||  | |:| | :
Db    373 SHGIYEGICNGVPMVMM-PLFGDQMDNAKRMETRGAGVTLNVLEMTADDLENALKTVINN 431

Qy    336 TNGNEQLTNMISTMEQDK 353
            :  | :   :|:: :|:
Db    432 KSYKENIMR-LSSLHKDR 448


RESULT   4
PCT-US92-00282-7
; Sequence 7, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  529 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US92-00282-7


  Query Match             5.1%;  Score 104.5;  DB 5;  Length 529;
  Best Local Similarity   18.4%;  Pred. No. 0.04;
  Matches   82;  Conservative   82;  Mismatches  166;  Indels  115;  Gaps   20;

Qy      9 LIITGS-FGNGHMQVTQSIVNQLNDMNL-DHLSVIEHDLFM---EAHPILTSICKKWYIN 63
            |:: ||  |:  : | |   : |:   : :|||   ||: :   | : :|     |:|
Db     17 LVLWGSVLGDKLLVVPQDGSHWLSMKEIVEHLSERGHDIVVLVPEVNLLLGE--SKYYRR 74

Qy     64 -------SFKYFRNMYKGF----------------YYSRPDKLDKCFYKYYGLNK---L 96
                : :  | |: |                  | :    :| ||:    | |
Db     75 KSFPVPYNLEELRTRYRSFGNNHFAASSPLMAPLREYRNNMIVIDMCFFSCQSLLKDSAT 134

Qy     97 INLLIKEKPDLILLTFPTPVMSVLTEQFNI----------------------NIP--- 129
          :: | : : | :       |   :| |   :                     :|
Db    135 LSFLRENQFDALFTDPAMPCGVILAEYLKLPSIYLFRGFPCSLEHIGQSPSPVSYVPRFY 194

Qy    130 ----------------VATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVK 173
                         :| :: :|  |      ||    :|:   |:|     :  :::
Db    195 TKFSDHMTFPQRLANFIANILENYLYH----CLYSKYEILASDLLKRDVSLPALHQNSLW 250

Qy    174 VTGIPIDNKFETPINQKQWLI-------DNNLDPDKQTILMSAGAFGV---SKGFDTMIT 223
          :       ::  |:       |          ||  : :  : ::|  |:   | |  :|::
Db    251 LLRYDFVFEYPRPVMPNMIFIGGTNCKKKGNLSQEFEAYVNASGEHGIVVFSLG--SMVS 308

Qy    224 DILAKSANAQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLM------ 277
          :|  | |  ::    |:   : ::|  ::  ||   :   |  : :|:  : |:
Db    309 EIPEKKA-MEIAEALGR---IPQTLLWRYTGTRPSNLAKNT-ILVKWLPQNDLLGHPKAR 363

Qy    278 --ITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFG-------KIADTPEEA 328
            ||  |   | ||    :||: : |  | :::||   | :| |          ||  | |
Db    364 AFITHSGSHGIYEGICNGVPMVMM-PLFGDQMDNAKRMETRGAGVTLNVLEMTADDLENA 422

Qy    329 IKIVASLTNGNEQLTNMISTMEQDK 353
          :| | :  :  | :   :|:: :|:
Db    423 LKTVINNKSYKENIMR-LSSLHKDR 446


RESULT   5
US-08-457-176-2
; Sequence 2, Application US/08457176
; Patent No. 5591826
;  GENERAL INFORMATION:
;    APPLICANT:  Vogelstein, Bert
;    APPLICANT:  Kinzler, Kenneth W.
;    APPLICANT:  de la Chappelle, Albert
;    TITLE OF INVENTION:  Mutator Gene and Hereditary
;    TITLE OF INVENTION:  No. 5591826-Polyposis Colorectal Cancer
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie, and Beckett
;      STREET:  1001 G Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,176
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/160295
;      FILING DATE:  02-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A.
;      REGISTRATION NUMBER:  32,141
;      REFERENCE/DOCKET NUMBER:  01107.44900
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202.508.9100
;      TELEFAX:  202.508.9299
;      TELEX:  197430 BBMB UT
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  934 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-457-176-2


  Query Match             5.0%;  Score 102.5;  DB 1;  Length 934;
  Best Local Similarity   22.3%;  Pred. No. 0.15;
  Matches   77;  Conservative   45;  Mismatches  131;  Indels   93;  Gaps   17;

Qy      18 GHMQVTQSIVNQLNDMNL--DHLSVIE------------HDLFMEAHPILTSICKKWYI 62
           |   | | |   | | |   : |:::|             || :   | |  : ||
Db     338 GQRLVNQWIKQPLMDKNRIEERLNLVEAFVEDAELRQTLQEDL-LRRFPDLNRLAKK--- 393

Qy      63 NSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLL-IKEKPD-----LILLTFPTPV 116
                     :  :   |  |:  | |:|:| |::   || :     |:|  | ||:
Db     394 ------------FQRQAANLQDCYRLYQGINQLPNVIQALEKHEGKHQKLLLAVFVTPL 440

Qy     117 MSVLTE--QFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKV 174
           : ::  :|   |      |     |:  : |                    ||:  ::
Db     441 TDLRSDFSKFQEMIETTLDMDQVENHEFLVKP------------------SFDPNLSEL 481

Qy     175 TGI--PIDNKFE-TPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSAN 231
           |   :: | : | |:  :   |  ||| ||  | |:  ||            :|  : |
Db     482 REIMNDLEKKMQSTLISAAR---DLGLDPGKQIKLDSSAQFGYYFRVTCKEEKVLRNNKN 538

Qy     232 AQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITITEGFA 291
           | |     :   |     |||   |   |||:  |:  :   :: :   : |: |:
Db     539 FSTVDIQKNGVKFTNS-----KLTS--LNEEYTKNKTEYEEAQDAIVKEI--VNISSGYV 589

Qy     292 RCIPMIFLNPAPGQELENAFYFE---------------EKGFGKI 321
           ||  ||    | |:   |                  ||| |:|
Db     590 E--PMQTLNDVLAQ-LDAVVSFAHVSNGAPVPYVRPAILEKGQGRI 632


RESULT   6
US-08-457-175-2
; Sequence 2, Application US/08457175
; Patent No. 5693470
;  GENERAL INFORMATION:
;    APPLICANT:  Vogelstein, Bert
;    APPLICANT:  Kinzler, Kenneth W.
;    APPLICANT:  de la Chappelle, Albert
;    TITLE OF INVENTION:  Mutator Gene and Hereditary
;    TITLE OF INVENTION:  No. 5693470-Polyposis Colorectal Cancer
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner, Birch, McKie, and Beckett
;      STREET:  1001 G Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20001
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,175
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/160295
;      FILING DATE:  02-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A.
;      REGISTRATION NUMBER:  32,141
;      REFERENCE/DOCKET NUMBER:  01107.44900
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202.508.9100
;      TELEFAX:  202.508.9299
;      TELEX:  197430 BBMB UT
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  934 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-457-175-2


  Query Match             5.0%;  Score 102.5;  DB 1;  Length 934;
  Best Local Similarity   22.3%;  Pred. No. 0.15;
  Matches   77;  Conservative   45;  Mismatches  131;  Indels   93;  Gaps   17;

Qy      18 GHMQVTQSIVNQLNDMNL--DHLSVIE------------HDLFMEAHPILTSICKKWYI 62
           |   | | |   | | |   : |:::|             || :   | |  : ||
Db     338 GQRLVNQWIKQPLMDKNRIEERLNLVEAFVEDAELRQTLQEDL-LRRFPDLNRLAKK--- 393

Qy      63 NSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLL-IKEKPD-----LILLTFPTPV 116
                     :  :   |  |:  | |:|:| |::   || :     |:|  | ||:
Db     394 ------------FQRQAANLQDCYRLYQGINQLPNVIQALEKHEGKHQKLLLAVFVTPL 440

Qy     117 MSVLTE--QFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKV 174
           : ::  :|   |      |     |:  : |                    ||:  ::
Db     441 TDLRSDFSKFQEMIETTLDMDQVENHEFLVKP------------------SFDPNLSEL 481

Qy     175 TGI--PIDNKFE-TPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSAN 231
           |   :: | : | |:  :   |  ||| ||  | |:  ||            :|  : |
Db     482 REIMNDLEKKMQSTLISAAR---DLGLDPGKQIKLDSSAQFGYYFRVTCKEEKVLRNNKN 538

Qy     232 AQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITITEGFA 291
           | |     :   |     |||   |   |||:  |:  :   :: :   : |: |:
Db     539 FSTVDIQKNGVKFTNS-----KLTS--LNEEYTKNKTEYEEAQDAIVKEI--VNISSGYV 589

Qy     292 RCIPMIFLNPAPGQELENAFYFE---------------EKGFGKI 321
           ||  ||    | |:   |                  ||| |:|
Db     590 E--PMQTLNDVLAQ-LDAVVSFAHVSNGAPVPYVRPAILEKGQGRI 632


RESULT   7
US-08-709-784-1
; Sequence 1, Application US/08709784
; Patent No. 6048701
;  GENERAL INFORMATION:
;    APPLICANT:  The Johns Hopkins University
;    TITLE OF INVENTION:  Antibody Detection of Mismatch Repair
;    TITLE OF INVENTION:  Protein
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Banner & Allegretti, Ltd.
;      STREET:  1001 G Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20001-4597
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/709,784
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/480,351
;      FILING DATE:
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kagan, Sarah A.
;      REGISTRATION NUMBER:  32,141
;      REFERENCE/DOCKET NUMBER:  1107.57434
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-508-9100
;      TELEFAX:  202-508-9299
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  934 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-709-784-1


  Query Match             5.0%;  Score 102.5;  DB 3;  Length 934;
  Best Local Similarity   22.3%;  Pred. No. 0.15;
  Matches   77;  Conservative   45;  Mismatches  131;  Indels   93;  Gaps   17;

Qy      18 GHMQVTQSIVNQLNDMNL--DHLSVIE-------------HDLFMEAHPILTSICKKWYI 62
           |   | | |   | | |   : |:::|               || :   | |  : ||
Db     338 GQRLVNQWIKQPLMDKNRIEERLNLVEAFVEDAELRQTLQEDL-LRRFPDLNRLAKK--- 393

Qy      63 NSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLL-IKEKPD-----LILLTFPTPV 116
                 :  :   |  |:  | |:|:| |::   || :     |:|  | ||:
Db     394 ------------FQRQAANLQDCYRLYQGINQLPNVIQALEKHEGKHQKLLLAVFVTPL 440

Qy     117 MSVLTE--QFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKV 174
             : ::  :|   |     |     |:  : |                     ||:  ::
Db     441 TDLRSDFSKFQEMIETTLDMDQVENHEFLVKP------------------SFDPNLSEL 481

Qy     175 TGI--PIDNKFE-TPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILAKSAN 231
             |  .:: | : | |:  :   |  ||| ||  | |:  ||           :|  : |
Db     482 REIMNDLEKKMQSTLISAAR---DLGLDPGKQIKLDSSAQFGYYFRVTCKEEKVLRNNKN 538

Qy     232 AQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLMITKPGGITITEGFA 291
             | |     :   |     |||   |   |||:  |:  :   :: :   : |: |:
Db     539 FSTVDIQKNGVKFTNS-----KLTS--LNEEYTKNKTEYEEAQDAIVKEI--VNISSGYV 589

Qy     292 RCIPMIFLNPAPGQELENAFYFE---------------EKGFGKI 321
             ||  ||    | |:    |                  ||| |:|
Db     590 E--PMQTLNDVLAQ-LDAVVSFAHVSNGAPVPYVRPAILEKGQGRI 632


RESULT   8
US-08-846-762-85
; Sequence 85, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
;  APPLICANT: Lam, Joseph S.
;  APPLICANT: Burrows, Lori
;  APPLICANT: Charter, Deborah
;  APPLICANT: de Kievit, Teresa
;  TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
;  TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa
;  FILE REFERENCE: 6580-089
;  CURRENT APPLICATION NUMBER: US/08/846,762A
;  CURRENT FILING DATE: 1997-04-30
;  NUMBER OF SEQ ID NOS: 100
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 85
;   LENGTH: 376
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-08-846-762-85


  Query Match             5.0%;  Score 101.5;  DB 2;  Length 376;
  Best Local Similarity   22.4%;  Pred. No. 0.047;
  Matches   53;  Conservative   42;  Mismatches   79;  Indels   63;  Gaps   12;

Qy      93 LNKLINLLIKEKPDLILL-TFPTPVMSVLTEQFNINIPVATVMTDYR---LHKNW----- 143
           |  |  :| : |||::|:    |  ::     |  |||  |    |   |:  |
Db      75 LEGLKPILAEFKPDVVLVHGDTTTTLATSLAAFYQRIPVGHVEAGLRTGDLYSPWPEEAN 134

Qy     144 --ITPYSTRYYVATKET-KQDFIDVGIDPSTVKVTGIPIDNKFETPINQKQWLIDN---- 196
             :| :   |: :  || :|: :   :  | : :||        | |:   |: |
Db     135 RTLTGHLAMYHFSPTETSRQNLLRENVADSRIFITG-------NTVIDALLWVRDQVMSS 187

Qy     197 ------------NLDPDKQTILMSA---GAFGVSKGFDTM---ITDILAKSANAQVVMIC 238
                       :||||: ||::     :||  :||: :   : ||     : |:|
Db     188 DKLRSELAANYPFIDPDKKMILVTGHRRESFG--RGFEEICHALADIATTHQDIQIVYPV 245

Qy     239 GKSKELKRSLTAKFKLTRMYLILGYTKHM-----NEW------MASSQLMITKPGGI 284
             :  ::  :             |||: |::      |:      |  : |::|  |||
Db     246 HLNPNVREPVNR---------ILGHVKNVILIDPQEYLPFVWLMNHAWLILTDSGGI 293


RESULT   9
PCT-US92-00282-3
; Sequence 3, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  533 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US92-00282-3


  Query Match             4.9%;  Score 99.5;  DB 5;  Length 533;
  Best Local Similarity   19.3%;  Pred. No. 0.13;
  Matches   74;  Conservative   74;  Mismatches  132;  Indels  103;  Gaps   20;

Qy      16 GNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSFKYFRNMYKGF 75
             |  |:   : ::  | : : |   |:  | |:   ||:       |::    |      |
Db     126 GCSHLLHNKELMASLAESSFD---VMLTDPFLPCSPIVAQ-----YLSLPTVF------F 171

Qy      76 YYSRPDXLD----KC--FYKYYG------------LNKLINLLIKEKPDLILLTFPTPVM 117
```

```
Db    172 LHALPCSLEFEATQCPNPFSYVPRPLSSHSDHMTFLQRVKNMLIAFSQNFLCDVVYSPYA 231

Qy    118 SVLTEQFNINIPVATVMTDYRLHKNWITPYSTRYYVATKETKQDFI---DVGIDPSTVKV 174

Db    232 TLASEFLQREVTVQDLLSSASV---WL------------FRSDFVKDYPRPIMPNMVFV 275

Qy    175 TGI------PIDNKFETPINQKQWLIDNNLDPDKQTILMSAGAFGV---SKGFDTMITDI 225

Db    276 GGINCLHQNPLSQEFEAYIN------------------ASGEHGIVVFSLG--SMVSEI 314

Qy    226 LAKSANAQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLM-------- 277

Db    315 PEKKAMA-IADALGKNPQ---TVLWRYTGTRPSNLANNT-ILVKWLPQNDLLGHPMTRAF 369

Qy    278 ITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFG-------KIADTPEEAIK 330

Db    370 ITHAGSHGVYESICNGVPMVMM-PLFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALK 428

Qy    331 IVASLTNGNEQLTNMISTMEQDK 353

Db    429 AVINDKSYKENIMR-LSSLHKDR 450


RESULT 10
US-08-480-604A-10
; Sequence 10, Application US/08480604A
; Patent No. 5736139
;  GENERAL INFORMATION:
;    APPLICANT:  KINK, JOHN A.
;    APPLICANT:  THALLEY, BRUCE S.
;    APPLICANT:  PADHYE, NISHA V.
;    APPLICANT:  FIRCA, JOSEPH R.
;    APPLICANT:  STAFFORD, DOUGLAS C.
;    TITLE OF INVENTION:  VACCINE AND ANTITOXIN FOR TREATMENT AND
;    TITLE OF INVENTION:  PREVENTION OF C. DIFFICILE DISEASE
;    NUMBER OF SEQUENCES:  32
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/480,604A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/422,711
;      FILING DATE:  14-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/405,496
;      FILING DATE:  16-MAR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01763
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2366 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-480-604A-10


  Query Match             4.8%;  Score 98.5;  DB 1;  Length 2366;
  Best Local Similarity   20.0%;  Pred. No. 1.7;
  Matches   89;  Conservative   73;  Mismatches  129;  Indels  155;  Gaps   24;

Qy      2 VTQNKKIL---------IITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPI 52

Db   1377 IEENKIILNSHEINFSGEVNGS--NGFVSLTFSILEGIN-------AIIEVDLLSKSYKL 1427

Qy     53 LTS------------ICKKWYINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLIN-- 98

Db   1428 LISGELKILMLNSNHIQQK--IDYIGFNSELQKNIPYSFVDSEGK-------ENGFINGS 1478

Qy     99 ----LLIKEKPDLILLTFPTPVMSVLTEQ-------FNINIPVATVMTDYRLHKNWITPY 147

Db   1479 TKEGLFVSELPDVVLIS------KVYMDDSKPSFGYYSNNLKDVKVITKDNVN------I 1526

Qy    148 STRYYVATKETKQDFIDVGI-----DPSTVKVTGIPIDNKFETPINQKQWLIDNNLDPDK 202

Db   1527 LTGYYL------KDDIKISLSLTLQDEKTIKLNSVHLD---ESGVAE------------- 1564

Qy    203 QTILMSAGAFGVSKGFDTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLT------- 255

Db   1565 --ILKFMNRKGNTNTSDSLMS--FLESMNIKSIFVNFLQSNIKFILDANFIISGTTSIGQ 1620

Qy    256 -----------RMYLILGYTKHMNE--WMASSQLMITKP-------GGITIT-------- 287

Db   1621 FEFICDENDNIQPYFIKFNTLETNYTLYVGNRQNMIVEPNYDLDDSGDISSTVINFSQKY 1680

Qy    288 -EGFARCIPMIFLNP---------APGQELENAF---------YFEEKGFGKIADTPEEA 328

Db   1681 LYGIDSCVNKVVISPNIYTDEINITPVYETNNTYPEVIVLDANYINEKINVNINDL---S 1737

Qy    329 IKIVASLTNGNEQLTNMISTMEQDKI 354

Db   1738 IRYVWS-NDGNDFI--LMSTSEENKV 1760


RESULT 11
US-08-405-496A-10
; Sequence 10, Application US/08405496A
; Patent No. 5919665
;  GENERAL INFORMATION:
;    APPLICANT:  WILLIAMS, JAMES A.
;    TITLE OF INVENTION:  VACCINE FOR CLOSTRIDIUM BOTULINUM
;    TITLE OF INVENTION:  NEUROTOXIN
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MEDLEN & CARROLL, LLP
;      STREET:  220 MONTGOMERY STREET, SUITE 2200
;      CITY:  SAN FRANCISCO
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  US/08/405,496A
;      FILING DATE:  16-MAR-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/329,154
;      FILING DATE:  25-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/161,907
;      FILING DATE:  02-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/985,321
;      FILING DATE:  04-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/429,791
;      FILING DATE:  31-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  INGOLIA, DIANE E.
;      REGISTRATION NUMBER:  40,027
;      REFERENCE/DOCKET NUMBER:  OPHD-01308
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 705-8410
;      TELEFAX:  (415) 397-8338
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2366 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-405-496A-10


  Query Match             4.8%;  Score 98.5;  DB 2;  Length 2366;
  Best Local Similarity   20.0%;  Pred. No. 1.7;
  Matches   89;  Conservative   73;  Mismatches  129;  Indels  155;  Gaps   24;

Qy       2 VTQNKKIL---------IITGSFGNGHMQVTQSIVNQLNDMNLDHLSVIEHDLFMEAHPI 52
           : :|| ||          : ||  || : :| ||:  :|       ::|| ||  ::: :
Db    1377 IEENKIILNSHEINFSGEVNGS--NGFVSLTFSILEGIN-------AIIEVDLLSKSYKL 1427

Qy      53 LTS------------ICKKWYINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLIN-- 98
           | |            | :|  |:   :   : |   || |   |        |  ||
Db    1428 LISGELKILMLNSNHIQQK--IDYIGFNSELQKNIPYSFVDSEGK-------ENGFINGS 1478

Qy      99 ----LLIKEKPDLILLTFPTPVMSVLTEQ-------FNINIPVATVMTDYRLHKNWITPY 147
               | : | ||::|::       |  :         :: |:    |:|  . ::
Db    1479 TKEGLFVSELPDVVLIS------KVYMDDSKPSFGYYSNNLKDVKVITKDNVN------I 1526

Qy     148 STRYYVATKETKQDFIDVGI-----DPSTVKVTGIPIDNKFETPINQKQWLIDNNLDPDK 202
             | ||:      :| | : :     |  |:|:  : :|   |: : :
Db    1527 LTGYYL------KDDIKISLSLTLQDEKTIKLNSVHLD---ESGVAE------------ 1564

Qy     203 QTILMSAGAFGVSKGFDTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLT------- 255
             ||      | :   |::::    :| | : : :      :|  | | | ::
Db    1565 --ILKFMNRKGNTNTSDSLMS--FLESMNIKSIFVNFLQSNIKFILDANFIISGTTSIGQ 1620

Qy     256 -----------RMYLILGYTKHMNE--WMASSQLMITKP-------GGITIT-------- 287
                     : | |   |   |   :: : | || :|       | |: |
Db    1621 FEFICDENDNIQPYFIKFNTLETNYTLYVGNRQNMIVEPNYDLDDSGDISSTVINFSQKY 1680

Qy     288 -EGFARCIPMIFLNP---------APGQELENAF---------YFEEKGFGKIADTPEEA 328
            |   |:  : ::|           |  | | :          |  ||    | |    :
Db    1681 LYGIDSCVNKVVISPNIYTDEINITPVYETNNTYPEVIVLDANYINEKINVNINDL---S 1737

Qy     329 IKIVASLTNGNEQLTNMISTMEQDKI 354
           |: | |  :||: : ::|| |::|:
Db    1738 IRYVWS-NDGNDFI--LMSTSEENKV 1760


RESULT  12
PCT-US92-00282-5
; Sequence 5, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  531 amino acids
;      TYPE:  AMINO ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US92-00282-5


  Query Match             4.5%;  Score 93;  DB 5;  Length 531;
  Best Local Similarity   21.2%;  Pred. No. 0.58;
  Matches   85;  Conservative   64;  Mismatches  147;  Indels  104;  Gaps   22;

Qy      14 SFGNGHMQ----VTQSIVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWYINSFKYFR 69
           ||||  |      :|       |:| :  |  |     ::    |         | ||
Db      93 SFGNNHFAERSFLTAPQTEYRNNMIVIGLYFINCQSLLQDRDTL---------NFFK--- 140

Qy      70 NMYKGFYYSRPDKLDKCF-----------YKYYGLNKLINLLIKEKPDLILLTF---PTP 115
                   | |  |             :| ||  :   | :  |  :  ||   | |
Db     141 ----------ESKFDALFTDPALPCGVILAEYLGLPSV--YLFRGFPCSLEHTFSRSPDP 188

Qy     116 VMSV------LTEQFNINIPVATVMTDYRLHKNWITPY------STRYYVATKETKQDFI 163
           |  :       ::    :  ||  :        | : ||      |    :|:   |:| :
Db     189 VSYIPRCYTKFSDHMTFSQRVANFLV------NLLEPYLFYCLFSKYEKLASAVLKRD-V 241

Qy     164 DVGIDPSTVKVTGIPIDNKFETP-----------INQKQWLIDNNLDPDKQTILMSAGA 211
           |: |  | | |  :  |   | |             || |:     :|  : :  : ::|
Db     242 DI-ITLSEVSVWLLRYDFVLEYPRPVMPNMVFIGGINCKK---RKDLSQEFEAYINASGE 297

Qy     212 FGV---SKGFDTMITDILAKSANAQVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMN 268
            |:   | |  :|:::|  | | | :    ||: :   ::  ::  ||   :   |  :
Db     298 HGIVVFSLG--SMVSEIPEKKAMA-IADALGKNPQ---TVLWRYTGTRPSNLANNT-ILV 350

Qy     269 EWMASSQLM--------ITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFG- 319
           :|:  : |:         ||  |   : |     :||: : |  | :::||   | || |
Db     351 KWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMM-PLFGDQMDNAKRMETKGAGV 409

Qy     320 ------KIADTPEEAIKIVASLTNGNEQLTNMISTMEQDK 353
                 ::  | |:| | :   :  | :   :|:: :|:
Db     410 TLNVLEMTSEDLENALKAVINDKSYKENIMR-LSSLHKDR 448


RESULT  13
US-08-451-715A-4
```

```
; Sequence 4, Application US/08451715A
; Patent No. 5801013
;  GENERAL INFORMATION:
;    APPLICANT:  Tao, Jianshi
;    APPLICANT:  Qui, Yan
;    APPLICANT:  Houman, Fariba
;    APPLICANT:  Shen, Xiaoyu
;    APPLICANT:  Schimmel, Paul R.
;    TITLE OF INVENTION:  Helicobacter Aminoacyl-tRNA Synthetase
;    TITLE OF INVENTION:  Proteins, Nucleic Acids and Strains Comprising Same
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/451,715A
;      FILING DATE:  26-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E.
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  CPI94-25
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  648 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-451-715A-4


  Query Match             4.5%;  Score 92;  DB 1;  Length 648;
  Best Local Similarity   19.6%;  Pred. No. 0.99;
  Matches   80;  Conservative   58;  Mismatches  152;  Indels  118;  Gaps   19;

Qy       4 QNKKILIITGSFGNGHMQVTQS--IVNQLNDMNLDHLSVIEHDLFMEAHPILTSICKKWY 61
           | :::  :||:  :|  :: ||  : ||      | :| |  |             :|
Db      38 QGEEVFFLTGTDEHG-QKIEQSARLRNQSPKAYADSISAIFKD----------QW- 81

Qy      62 INSFKYFRNMYKGFYYSRPDKLDKC----------FYK----------YYGLNK 95
             :|   | ||  :   :  ||             ||             | ::|
Db      82 ----DFFNLDYDGFIRTTDSEHQKCVQNAFEIMFEKGDIYKGAYSGYYCVSCESYCAISK 137

Qy      96 LIN--------------LLIKEKPDLILLTFPTPVMSVLTEQFNINIPVATVMTDYRLH 140
             |              || :|     |  :  |::   :     :|:      ||
Db     138 ADNTSDKVLCPDCLRETALLEEESYFFKLSAYEKPLLEFYAKNPEAILPI------YR-- 189

Qy     141 KNWITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGIPIDNKFETPINQKQ-WLIDNNLD 199
           || :| :           :|  :|: |  :: :  |||:  |   | :    ||
Db     190 KNEVTSF----------IEQGLLDLSITRTSFE-WGIPLPKKMNDPKHVVYVWL------ 232

Qy     200 PDKQTILMSAGAFGVSKGFDTMITDILAKSANAQVVMICGK---SKELKRSLTAKFKLTR 256
             :|  | | |   | |          |      | :    |:: |     :
Db     233 ---DALLNYASALGYLNGLD----------NKMAHFECARHIVGKDILRFHAIYWPAFL 278

Qy     257 MYLILGYTKHM--NEWMASSQLMITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFE 314
           | | |   | :  : |     : ::|  |  :    |: : | :     | |    |
Db     279 MSLNLPLFKQLCVHGWWTIEGVKMSKSLGNVLD---AQKLAMEY-----GIEELRYFLLR 330

Qy     315 EKGFGKIADTPEEAI--KIVASLTNGNEQLTNMISTMEQDKIKYATQT 360
           |  ||:  |  ::|:  :| |:| |   | | :  | :    |: ::
```

```
Db     331 EVPFGQDGDFSKKALIERINANLNNDLGNLLNRLLGMAKKYFNYSLKS 378


RESULT  14
US-08-942-012B-32
; Sequence 32, Application US/08942012B
; Patent No. 6235278
; GENERAL INFORMATION:
;  APPLICANT:  Miller, Lois K.
;  APPLICANT:  Lu, Albert
;  APPLICANT:  Dierks, Peter
;  APPLICANT:  Black, Bruce
;  TITLE OF INVENTION:  Biological Insect Control  Agents Expressing
;  TITLE OF INVENTION:  Insect-Specific Toxin Genes, Methods and Compositions
;  FILE REFERENCE:  28-96a
;  CURRENT APPLICATION NUMBER: US/08/942,012B
;  CURRENT FILING DATE:  1997-10-01
;  PRIOR APPLICATION NUMBER:  08/729,606
;  PRIOR FILING DATE:  2000-10-01
;  NUMBER OF SEQ ID NOS:  33
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 32
;   LENGTH: 515
;   TYPE: PRT
;   ORGANISM: Spodoptera littoralis nuclear polyhedrosis virus
US-08-942-012B-32


  Query Match             4.4%;  Score 91;  DB 4;  Length 515;
  Best Local Similarity   24.0%;  Pred. No. 0.87;
  Matches   55;  Conservative   33;  Mismatches  101;  Indels   40;  Gaps   10;

Qy     169 PSTVK-VTGIPIDNKFETPINQKQWLIDNNLDPDKQTILMSAGAFGVSKGFDTMITDILA 227
           || |: : || ||    :  ::    |||:|    :   |  |   || |     :|: :
Db     265 PSNVQYLGGIHIDPAVTSVADE----IDNDLAEFLENSTM--GVVYVSLGSSVRASDMDS 318

Qy     228 KSANA----------QVVMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEWMASSQLM 277
             |              :|:    || ::  :: :   : | :      || |      ::
Db     319 NMLNVFVETFRSIPYRVLWKVDKSDKIFDNIPSNVLIQRWFPQRRVLKHRN-----VKVF 373

Qy     278 ITKPGGITITEGFARCIPMIFLNPAPGQELENAFYFEEKGFGKIADTPEEAIKIVASLTN 337
           ||: |  :  |     :|| |  |  |  :  | : :|  | |:  ||           ||
Db     374 ITQGGVQSTDEAIDAGVPM-FGVPIMGDQFYNVYMYETYGIGRGVDT----------LTV 422

Qy     338 GNEQLTNMISTMEQDKIKYATQTI-CRDLLDLIGHSSQPQEIYGKVPLY 385
             ||| ::  :  |  ||   |:  || :       ||      |   |
Db     423 DARQLTEIVMDV-ADNEKYKNGTLWLRDAI-----MDQPMRPLEKAVWY 465


RESULT  15
US-09-005-180A-4
; Sequence 4, Application US/09005180A
; Patent No. 6124446
;  GENERAL INFORMATION:
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  HUMAN VPS35/MEM3-RELATED PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Dr.
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/005,180A
```

```
;      FILING DATE:  Filed January 8, 1998
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0457 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  937 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  854543
US-09-005-180A-4


  Query Match             4.3%;  Score 89;  DB 4;  Length 937;
  Best Local Similarity   18.0%;  Pred. No. 3.5;
  Matches   72;  Conservative   66;  Mismatches  125;  Indels  136;  Gaps   18;

Qy       2 VTQNKKILIITGSFGNGHMQVTQSIVNQLN--------DMNL-----DHLSVIEH---DL 45
           |: || :| :      :    ||: : :  |        ||::     |:|:|: |   ||
Db     274 VSINKIVLTLVDRLND---YVTRQLEDDPNATSTNAYLDMDVFGTFWDYLTVLNHERPDL 330

Qy      46 FMEAH-PILTSI---CKKWYINSFKYFRNMYKGFYYSRPDKLDKCFYKYYGLNKLINLLI 101
            ::   |:: |:     ||| |:|                          ||||  |::
Db     331 SLQQFIPLVESVIVLSLKWYPNNFD-----------------------NLNKLFELVL 365

Qy     102 KEKPD-------------LILLTFPTPVMSVLTEQFNI--NIPVAT-------VMTDYR 138
           ::  |              |:||:|    :  ||       | || :       :::  :
Db     366 QKTKDYGQKNISLESEHLFLVLLSFQNSKLQ-LTSSTTAPPNSPVTSKKHFIFQLISQCQ 424

Qy     139 LHKNWITPYSTRYYVATKETKQDFIDVGIDPSTVKVTGIPIDNKFETPINQKQWLIDNNL 198
            :|| :   |       |:   : ||: :|    ::     ::|   | :    :|:: |
Db     425 AYKNILALQSIS---LQKKVVNEIIDILMDREVEEMADNDSESKLHPPGHSAYLVIEDKL 481

Qy     199 DPDK------QTILMSAGAFGVSKGFDTMITDI-------------------------- 225
              :       |:  :|        || : ::
Db     482 QVQRLLSICEPLIISRSGPPANVASSDTNVDEVFFNRHDEEESWILDPIQEKLAHLIHWI 541

Qy     226 ------------LAKSANAQV-VMICGKSKELKRSLTAKFKLTRMYLILGYTKHMNEW- 270
                       :  |  ||: :::  ||  :|  :  |:  |   :|  : | | :
Db     542 MNTTSRKQTMKNKIQFSLEAQLEILLLIKSSFIKGGINVKY--TFPAIITNFWKLMRKCR 599

Qy     271 MASSQLMITKPGGITITEGF-----------ARCIPMIF 298
           |    |:  :|   |:   :           :|||  ||
Db     600 MIQEYLLKKRPDNKTLLSHYSNLLKQMFKFVSRCINDIF 638


Search completed: June 29, 2001, 08:59:41
Job time: 179 sec
```